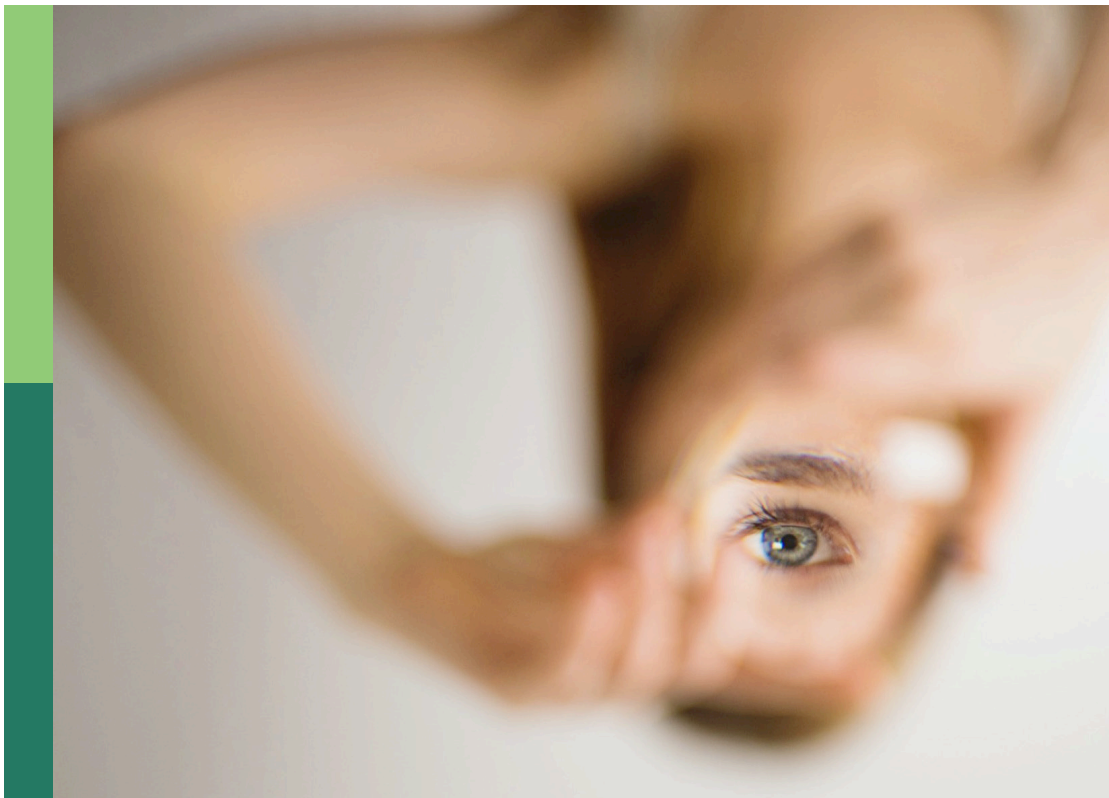
frontiers | Research Topics

Gender differences and disparities in socialization contexts: How do they matter for healthy relationships, wellbeing, and achievement-related outcomes?

Edited by
Caterina Fiorilli, Daniela Barni, Jan Retelsdorf and Joyce Endendijk

Published in
Frontiers in Psychology
Frontiers in Education



ISSN 1664-8714
ISBN 978-2-83251-076-6
DOI 10.3389/978-2-83251-076-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Gender differences and disparities in socialization contexts: How do they matter for healthy relationships, wellbeing, and achievement-related outcomes?

Topic editors

Caterina Fiorilli — Libera Università Maria SS. Assunta University, Italy
Daniela Barni — University of Bergamo, Italy
Jan Retelsdorf — University of Hamburg, Germany
Joyce Endendijk — Utrecht University, Netherlands

Citation

Fiorilli, C., Barni, D., Retelsdorf, J., Endendijk, J., eds. (2023). *Gender differences and disparities in socialization contexts: How do they matter for healthy relationships, wellbeing, and achievement-related outcomes?* Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83251-076-6

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 02 December 2022
DOI 10.3389/fpsyg.2022.1103425

OPEN ACCESS

EDITED AND REVIEWED BY
Claudio Longobardi,
University of Turin, Italy

*CORRESPONDENCE
Caterina Fiorilli
fiorilli@lumsa.it

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 20 November 2022
ACCEPTED 22 November 2022
PUBLISHED 02 December 2022

CITATION
Fiorilli C, Barni D, Endendijk J and Retelsdorf J (2022) Editorial: Gender differences and disparities in socialization contexts: How do they matter for healthy relationships, wellbeing, and achievement-related outcomes?
Front. Psychol. 13:1103425.
doi: 10.3389/fpsyg.2022.1103425

Editorial: Gender differences and disparities in socialization contexts: How do they matter for healthy relationships, wellbeing, and achievement-related outcomes?

Caterina Fiorilli[1]*, Daniela Barni[2], Joyce Endendijk[3] and Jan Retelsdorf[4]

[1]Department of Human Sciences, Libera Università Maria SS. Assunta (LUMSA), Rome, Italy, [2]Department of Human and Social Sciences, Università degli Studi di Bergamo, Bergamo, Italy, [3]Child and Adolescent Studies, Utrecht University, Utrecht, Netherlands, [4]Faculty of Education, Educational Psychology, University of Hamburg, Hamburg, Germany

KEYWORDS

gender differences, gender disparities, school, teachers, family, parents, youth, wellbeing

Editorial on the Research Topic
Gender differences and disparities in socialization contexts: How do they matter for healthy relationships, wellbeing, and achievement-related outcomes?

Gender differences and disparities in youth's development, education, and socialization are part of long-standing scientific, political, and public debates. According to the European Institute for Gender Equality (https://eige.europa.eu/), gender disparities refer to differences in women's and men's access to resources, status, and wellbeing, which usually favor men and not rarely are institutionalized through law, justice, and social norms. Despite remarkable advances in furthering the status of women, gender disparities still remain a worldwide challenge, as no country has achieved full gender parity yet (World Economic Forum, 2022). At the current rate of progress, it will take 132 years to close the global gender gap. Gender disparities largely persist in several life domains such as school (e.g., in academic pathway and achievement), work (e.g., career development and wages), and family (e.g., household division and parental expectations of children), and can result in context- and gender-specific problems and maladjustment. It is thus essential to better understand the psychosocial mechanisms underlying gender differences in socialization contexts in order to reduce the risk of harmful disparities and strengthen the factors fostering equitable development opportunities for girls and boys.

With a multiperspective approach, the current Research Topic (RT) aims to contribute to the international debate by offering scientific data and educational and

social suggestions for building a social context supporting optimal development of youth, regardless of their gender. The following sections describe the RT's contributions in two sub-themes.

Gender disparities: From school to university

Most current RT papers allow us to observe how the gender gap in the school context persists in many countries (Austria, Australia, Colombia, Denmark, Germany, Italy, Japan, Quebec, Nigeria, Switzerland, and United States) and at different school-ages. Together these studies highlight the need for extra attention to gender differences in the school context by education staff and policymakers.

A large body of literature is devoted to girls' and boys' attitudes and performance in science, technology, engineering, and mathematics (STEM). In line with Eccles and Wigfield's (2020) situated expectancy-value theory, many sociocultural, contextual, biological, behavioral, and psychological variables may contribute to the widespread under-representation of girls and women in the scientific field and a lower academic self-concept than boys. In this regard, Valls's research has confirmed gender differences in academic self-concept with girls feeling more confident in language learning and boys feeling more confident in mathematics. Furthermore, Valls's research demonstrated that negative social comparison processes could best explain these gender differences, which, in turn, may negatively impact boys' and girls' motivation toward certain academic challenges. Similarly, Andersen and Smith found that the social contexts in schools (i.e., teacher gender stereotypes, comparisons with math achievement of female peers) generates gender differences in young people's self-concept and achievements in math and language. In Hübner et al.'s study clear disparities favoring boys were found for upper secondary school achievements in math and physics and to a lesser extent in biology. These disparities did not increase (nor decrease) after a recent school time reform in Germany that reduced overall school time, which was compensated by increased average instructional time per week. Although, girls' level of stress and wellbeing was negatively affected by this instructional time reform to a greater extent than for boys, which may on a longer term exacerbate existing gender disparities in the school context.

Interestingly, as Froehlich et al. outlined, although there are no gender differences in math ability in young STEM students, expected backlash (i.e., less positive reactions to university major) affected female STEM students' emotions and STEM motivation to a larger extent than male STEM students. Despite the relatively higher level of female students' mathematics achievement than boys, they maintain a weaker math self-concept, negatively affecting the cognitive resources necessary to perform STEM tasks better (Bertrams et al.).

Similarly, Musso et al. focused on STEM-gender stereotypes and assumed that gender disparities become more complex and pronounced when socioeconomic status (SES) is considered. The authors shed light on the unneglectable consideration that higher SES is associated with lower STEM-gender stereotypes. With a different approach to SES, Kuzyk et al. confirmed the interrelationships between SES, nationality, and gender, which may interactively impact students' cognitive performance and self-perceptions of this performance. Additionally, despite evidence that IQ levels are equally distributed between genders, there is a significant gender gap in self-estimated intelligence, with males providing systematically higher estimates than females (Reilly et al.).

How gender-stereotypes and disparities threaten adolescents' mental health and wellbeing is a second Research Topic concerning gender disparities at school. According to Rubach et al., it is not a surprise that male and female students report distinct stressors and mental health troubles contextually observed during the COVID-19 pandemic. Nevertheless, teachers' instructional quality may reduce mental health menaces and enhance students' academic satisfaction. Similarly, Korlat et al. focused on gender role self-concept (i.e., masculine, feminine, androgynous, and undifferentiated) in relation to school-related wellbeing. Their findings showed that an androgynous self-concept might be optimal for academic wellbeing. Furthermore, their study opens urgent reflections on how school staff might approach gender-typed attributes in students.

With a different perspective on the educational setting, the third theme of the RT focuses on the relationships between teachers' gender and their mental health. Kreuzfeld and Seibt shed an interesting light on how male and female teachers differ in terms of working conditions and coping with high work demands, as well as individual factors that promote early retirement. By collecting several types of data from a gender-balanced group of teachers, the authors found that female teachers have a greater tendency to overcommit themselves and have a worse capacity to recover from troubles than male teachers. A second study by Dersch et al. addressed educators' stereotypes regarding STEM and outlined that teachers' misconceptions may impact their students' self-concepts. Preservice teachers' training should thus promote their awareness of gender misconceptions.

The focus on teacher-student relationships was also analyzed in the research by Beißert et al. concerning teachers' reactions to social exclusion among students by considering their gender. Interestingly, teachers were less likely to intervene if a boy was excluded than if a girl was excluded. This study drew attention to male-specific school disparities by showing that also boys can be at risk of being encapsulated in their gender role, which, in turn, may negatively affect their school-adjustment.

Finally, Bluteau et al. analyzed the relationship between students' seating in the classroom and their school-related wellbeing. Flexible classroom seating positively affects girls' wellbeing, while male students take advantage of fixed classroom seating. Thus, seating arrangements, and individual differences in the need for personal space, could contribute the gender gap in wellbeing at school.

An important future direction for research on gender disparities in the school context is to not primarily focus on gender in STEM, but also examine processes related to the underrepresentation of boys and men in HEED (health care, elementary education, domestic sphere; Croft et al., 2015) as well as gender differences in the performance on other school subjects (e.g., language, arts).

Reducing gender disparities: Start early, at home

For the greater part of childhood and early adolescence, the family is another primary context in which children and youths are socialized about gender and gender roles (e.g., Lawson et al., 2015). Parents engage in numerous *cultural socialization* processes and practices, which expose children to information that helps them to learn about their history, heritage (values, religion, traditions, customs, etc.), and social norms (e.g., what is socially expected from a girl or a boy). One such cultural process among families is parent–child transmission of norms, beliefs, and values which many scholars consider the hallmark of successful intergenerational socialization (Knafo-Noam et al., 2020). Parents widely use perceived social norms and stereotypical beliefs as a reference when socializing children (Tam et al., 2012). This clearly emerged from Barni et al.'s study, which showed a significant relationship between parents' hostile and benevolent sexism and their socialization values (i.e., the values parents want to transmit to their children). The more parents, especially fathers, hold sexist beliefs against women, the more they would like their young adult children to be conservative.

Parents' beliefs translate into daily practices and influence children's development of competencies and motivations. In this regard, Mues et al., involving preschool children, showed that parents' mathematical gender stereotypes (in favor of boys), self-efficacy, and their beliefs on the importance of mathematical activities at home are related to parents' numeracy activities and children's numeracy competencies. The findings supported the assumption of a direct association between children's numeracy competencies and parents' numeracy-related activities for fathers only, but not for mothers. In general, parents' gender-differentiated encouragement of science or language predicts children's later motivations (Shirefley and Leaper) and even career decisions (Endendijk and Portengen). Everhart Chaffee and Plante's results suggested that parents' ability stereotypes about language support girls' motivation for language arts; on the other hand, stereotypes that language arts are not for boys push them toward science. Boys are less interested in female-dominated fields, also regarding occupation, particularly when they feel pressure to conform to gender norms and hold stereotypical beliefs about these occupations (Masters and Barth). Endendijk and Portengen showed that parents' gender-typical career and family involvement (i.e., work hours and task division in the home) influence their children's vision of their future work and family roles. Children play an active role in developing this vision for the future through their gender identity, precisely by how similar they feel to individuals of the same gender.

Parental influence is so pervasive in children's acquisition of gender roles, knowledge, and understanding that perceived parenting styles are even related to young adults later intimate relationships outside the family. Paleari et al., in their study on cyber dating abuse, pointed out that the more young adults report that their mothers' parenting style was authoritarian or permissive during their childhood, the more likely they are to be involved in a cyber-abusive dating relationship. They have also found that mothers' parenting styles interact with fathers' styles in relating to their daughters' cyber control and aggression.

The studies included in this RT support the specific and interrelated role of fathers and mothers in children's gender socialization, substantially in the direction of conforming to gender stereotypes. In all these processes, children's sex and gender identity (Endendijk and Portegen) come into play by influencing parents' styles and practices and moderating their impact. Most gender disparities are harmful to girls at a young age, but some involve boys (see Everhart Chaffee and Plante), and they have long-term effects on academic paths, careers, and intimate relationships. It is nevertheless worthwhile noting that, under some individual and/or contextual conditions, the family can actively counteract cultural stereotypes about gender. For example, Shirefley and Leaper reported that highly educated parents—living near scientific/technology industries where women are employed—tend to use a higher proportion of science talk with daughters compared to sons.

These findings highlight that the psychosocial and educational programs to reduce the gender gap should start early at home by involving both parents. They could help parents to become more aware of their own gender-based biases and gender socialization practices, especially when these negatively impact children's health, by generating disparities (in terms of effective and symbolic opportunities), compromising children's

(eudaimonic) wellbeing, and feeding feelings of unfairness across generations.

Conclusion

Bringing together the above contributions, a multisystemic view of gender issues arises where different microsystems (mainly school and family) and sometime mesosystems (i.e., interactions across the microsystems) and macrosystems (i.e., cultures) are considered. This view can help in expanding focus to tap into a more comprehensive picture of gender differences and disparities and their consequences on youth's wellbeing in multiple daily life contexts so to inform social policies, provide intervention targets, and create a new community awareness of the roots of gender inequalities in current society.

Almost all the studies included in this RT provide a binary classification for gender. It would be worthwhile that future contributions on gender disparities in school and family contexts move beyond the binary toward a more multidimensional view of gender.

Author contributions

CF and DB wrote the draft of this editorial. JE and JR contributed to revise it. All authors have read and approved the final manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Croft, A., Schmader, T., and Block, K. (2015). An underexamined inequality: cultural and psychological barriers to men's engagement with communal roles. *Personal. Social Psychol. Rev.* 19, 343–370. doi: 10.1177/1088868314564789

Eccles, J. S., and Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. *Contemp. Educ. Psychol.* 61, 101859. doi: 10.1016/j.cedpsych.2020.101859

Knafo-Noam, A., Barni, D., and Schwartz, S. (2020). "Parent-child value similarity: broadening from intergenerational transmission to reciprocal influences, genetics, and environmental antecedents," in *The Oxford Handbook of Moral Development: An Interdisciplinary Perspective,* ed. L. Arnett Jensen, 164–185. Oxford: Academic.

Lawson, K. M., Crouter, A. C., and McHale, S. M. (2015). Links between family gender socialization experiences in childhood and gendered occupational attainment in young adulthood. *J. Voc. Behav.* 90, 26–35. doi: 10.1016/j.jvb.2015.07.003

Tam, K.-P., Lee, S.-L., Kim, Y.-H., Li, Y., and Chao, M. M. (2012). Intersubjective model of value transmission: parents using perceived norms as reference when socializing children. *Personal. Social Psychol. Bull.* 38, 1041–1052. doi: 10.1177/01461672124 43896

World Economic Forum (2022). *The Global Gender Gap Report 2022*. Geneva: The World Economic Forum Publishing. Available online at: http://reports.weforum.org/globalgender-gap-report-2022 (accessed November 10, 2022).

frontiers in Education

BRIEF RESEARCH REPORT
published: 11 January 2022
doi: 10.3389/feduc.2021.815619

Gender Differences in Social Comparison Processes and Self-Concept Among Students

*Marjorie Valls**

UER Développement de l'enfant à l'adulte, Haute Ecole Pédagogique Vaud, Lausanne, Switzerland

Forced social comparison (i.e., comparing oneself to another "predefined" student) has often been studied in school settings. However, to our knowledge, studies that explore its association with academic self-concept have rarely distinguished between subjects involved (e.g., mathematics or language learning). Moreover, some processes taking place during forced social comparison are thought to have a negative impact on academic self-concept. Thus, the aims of this study were to explore: 1) the associations between self-concepts (i.e., Language learning, Mathematics and Social), attitudes towards school and social comparison processes in school settings; and 2) the influence of social comparison processes on components of academic self-concept across gender. A sample of 238 elementary school students (M_{age} = 10.12, SD = 1.25; 52% boys) completed a questionnaire assessing self-concepts and attitudes towards school, as well as a questionnaire measuring four social comparison processes. Results indicated that girls used negative processes (i.e., upward contrast and downward identification) more than boys. In addition, boys reported better self-concept in mathematics while girls reported better self-concept in language learning (small effect). Results of stepwise multiple linear regression analyses showed that upward contrast best explained gender differences, with a stronger effect for girls. Attitudes towards school only explained gender differences in language learning self-concept. Furthermore, positive processes (i.e., upward identification and downward contrast) have no effect on either component of academic self-concept. Results of this study demonstrate the need to examine the evolution of social comparison processes over time, considering their impact on students' academic/social well-being and achievement from a gender perspective.

Keywords: forced social comparison, upward comparisons, downward comparisons, academic self-concept, gender, school settings

OPEN ACCESS

Edited by:
Caterina Fiorilli,
Libera Università Maria SS. Assunta University, Italy

Reviewed by:
Giacomo Angelini,
Libera Università Maria SS. Assunta University, Italy
Sonia Mangialavori,
University of Milan, Italy

***Correspondence:**
Marjorie Valls
marjorie.valls@hepl.ch

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *15 November 2021*
Accepted: *21 December 2021*
Published: *11 January 2022*

Citation:
Valls M (2022) Gender Differences in Social Comparison Processes and Self-Concept Among Students. Front. Educ. 6:815619. doi: 10.3389/feduc.2021.815619

INTRODUCTION

Social comparison theory has been applied to many clinical problems, such as body image, depression and burnout (Dijkstra et al., 2010). Social comparison includes all processes aimed at comparing one's own personal characteristics with those of others (Buunk and Gibbons, 2000). Dijkstra et al. (2010, p. 196) have identified processes involved when individuals compare themselves to others. According to them, individuals will choose various comparison targets. It refers to the direction of comparison: either a comparison with people judged as having similar abilities to their own (i.e., lateral comparisons), or a comparison with people having superior abilities (i.e., upward

comparisons) or inferior abilities (i.e., downward comparisons). Moreover, individuals would compare themselves in a movement of identification/contrast with respect to the chosen target. It refers to the framing of comparison: they can either identify with the comparison target by focusing on their similarities, or contrast themselves from the comparison target by focusing on their differences. Thus, four social comparison processes have been highlighted: upward identification, downward contrast, upward contrast and downward identification (Smith, 2000; Buunk et al., 2005). These four processes will be the point of reference for this article because they have already been studied in elementary school students (Boissicat et al., 2012; Bouffard et al., 2014). According to the meta-analysis by Gerber et al. (2018), contrast would be the dominant response as identification would require a special priming. Although the tendency would be to evaluate oneself positively, these authors believe that individuals "look upward to confirm their closeness to the 'better ones,' which often leads, alas, to self-deflation" (p. 194).

Specifically in school settings, social comparison can be defined as a student taking one or more classmates as comparison target in order to conduct an assessment of his/her own competence (Bouffard et al., 2014). Therefore, it would impact students' self-concept, especially academic self-concept. Internal/External frame of reference model (I/E model; Marsh, 1986) assumes the influence of social comparison on academic self-concept (Wolff et al., 2018). The internal frame of reference involves an internal comparison called "dimensional comparison" (e.g., student comparing his/her competence between two different subjects) while the external frame of reference involves an external comparison called "social comparison" (Ertl et al., 2017; Wolff et al., 2018). Wolff et al. (2018) indicate that if the student compares his/her academic performance to that of his/her peers and believes that he/she is better than them, this social comparison should lead to a better academic self-concept. Although the I/E model suggests that social and dimensional comparisons are jointly involved during the development of self-concept in school settings (Wolff et al., 2018), only the external frame of reference will be considered in this study. Indeed, the comparisons students make within their classrooms provide an external frame of reference for self-assessment and performance attribution (Ertl et al., 2017; see also; Rost et al., 2005) and appear to be most important when students form their academic self-concept (Wolff et al., 2018).

Boissicat et al. (2020) point out that when a student compares himself/herself to a classmate, this individual comparison may be unconscious (i.e., not being fully aware that he/she is comparing oneself), deliberate or forced. These authors define deliberate comparison as being fully chosen by students, by voluntarily selecting a comparison target within the classroom. It is assessed through nomination. Forced comparison, on the other hand, occurs when students are asked to compare themselves to another "predefined" student. In this particular case, students would appear to favor upward identification and downward contrast, both of which are non-threatening processes to the self (Bouffard et al., 2014). However, social comparison in school settings has often been defined as a type of vicarious experience, where observing a peer of the same level succeeding or failing in a task would provide information leading students to believe that they are likely to do the same (e.g., Boissicat et al., 2020). Nevertheless, according to Dijkstra et al. (2008) "although the concepts of modeling and social comparison overlap, they differ significantly" (p. 841). They state that the purpose of modeling is observation and imitation teaching (e.g., a procedure) that would ensure student success and thus positively influence academic self-concept. Conversely, social comparison occurs when students choose a target (i.e., another student) with whom to compare his/her competence or performance. Thus, "upward comparisons negatively affect students' academic self-concept" (Dijkstra et al., 2008, p. 841).

Positive effects of upward identification have been suggested in studies evaluating forced social comparison processes. Indeed, the preferred process for elementary school students appears to be upward identification followed by downward contrast similarly across gender, which are positively associated with perceived academic competence (Boissicat et al., 2012; Bouffard et al., 2014). Bouffard et al. (2014) indicate that this preference may be related to a more pronounced search for positive emotions that these two processes are presumed to generate. Nevertheless, Boissicat et al. (2012) found that upward identification would have a low contribution to academic self-concept, while downward identification would have the largest negative contribution despite its low use by students. These deleterious effects would be found even after controlling for academic performance. These authors conclude that links between self-concept and social comparison in school settings would not only depend on the direction of the comparison, but also on the framing. In addition, Dumas and Huguet (2011) point out that upward identification would be more implemented during deliberate comparisons, especially from the age of 10, with an effect of enhancing perceived competence. They also indicate that if the student is confronted with a failure, he/she will tend to take as comparison target a student with competence judged inferior to his/her own. Dumas and Huguet (2011) conclude that during forced comparisons (notably imposed by selective educational systems), positive effects of upward identification would not be sufficient to counteract the effects of upward contrast that generate a decline in academic self-concept.

Unfortunately, to our knowledge, studies that consider associations between academic self-concept and forced social comparison make little or no distinction between the subjects concerned. Moreover, depending on the age groups observed, results relating to gender differences seem to differ. For example, Pulford et al. (2018) showed that female university students were more likely to use upward comparisons, while male students were more likely to use downward comparisons. In addition, downward comparisons would not be related to academic confidence. Studies among elementary school students generally show that girls use downward identification and upward contrast more than boys (Boissicat et al., 2012; Bouffard et al., 2014). However, these two studies do not explore the relative contribution of these four social comparison processes to academic self-concept across subjects

and gender. Thus, the present study has a twofold purpose. First, it aims at exploring gender differences in academic self-concept in two important subjects during elementary grades (i.e., language learning and mathematics), social self-concept, attitudes towards school, and four social comparison processes previously identified. Second, it aims at exploring the influence of social comparison processes on academic self-concept in each subject area and by gender, taking into account effects of social self-concept and attitudes towards school. In particular, given the above-mentioned elements and age of the students involved in this study, we expect that girls will report implementing negative processes more frequently, which will have a negative impact on their academic self-concept. Furthermore, since the presumed positive effects of upward identification would not be sufficient in forced social comparison contexts (Dumas and Huguet, 2011), we expect that this process would not emerge as a significant predictor of academic self-concept.

MATERIALS AND METHODS

Research Design and Study Procedure

Data were collected from elementary school students (five to height graders) in a French-speaking canton of Switzerland during the 2017 to 2020 school years. In Switzerland, grades five to height correspond to the fifth to eighth years of compulsory schooling. The Cantonal Commission on Ethics in Human Research (CER-VD)[1] provides authorizations for clinical trials and human research projects that fall within the application field of the Human Research Act (HRA)[2]. However, the HRA does not apply to research conducted on health-related data that has been collected anonymously or anonymized. The Cantonal data protection acts[3] concern personal and identifiable data: henceforth data is anonymized, it is no longer covered by the Act. This research was conducted in accordance with the Code of Research Ethics for the Universities of Teacher Education (CDHEP)[4] and the International Ethical Guidelines for Health-related Research Involving Humans[5]. In particular, the duty to inform was respected. Parents were informed by letters of the general objectives of the study, and could decline their child's participation in the data collection. The letters also contained the identity of the supervisor and the institution for which he or she worked, as well as a contact address. Students were also given the option to decline to participate in the study, as their participation was voluntary. Under these conditions, no refusals were recorded (i.e., return rate of 100%) and the anonymity of the participants was preserved.

Data were collected in the classroom by Bachelor students who administrated the questionnaires anonymously. Each item was read a first time to ensure understanding by students, and completing the questionnaires took between 15 and 30 min.

Participants and Study Procedure

The sample consisted of 238 students (age range: 8–13 years), including 114 girls (48% of the total sample; Mage = 10.18, SD = 1.31) and 124 boys (Mage = 10.07, SD = 1.19). Of the total sample, 21.80% were fifth graders (21.90% girls; 21.80% boys), 23.90% were sixth graders (21.90% girls; 25.80% boys), 29.40% were seventh graders (29.80% girls; 29.00% boys) and 24.80% were eighth graders (26.30% girls; 23.40% boys).

Measures

Social Comparison Processes in School Context

Social comparison processes were assessed using the French version of the Questionnaire of the comparison of academic self (Questionnaire de la Comparaison de Soi Scolaire; QCSS) developed by Bouffard et al. (2014). The QCSS is a self-report questionnaire designed to assess four social comparison processes (i.e., upward contrast, α = 0.73; upward identification, α = 0.60; downward identification, α = 0.78; downward contrast, α = 0.77). Each process is composed of 3 items scored on a 4-point Likert scale (1 = "Not at all like me" to 4 = "Totally like me"; scores per dimension range from 3 to 12). A higher score on one dimension indicates higher frequency of use of a forced social comparison process. Confirmatory Factor Analysis (CFA) results show a good fit to the data (χ^2/df = 1.74; CFI = 0.96; RMSEA = 0.06; SRMR = 0.05). With regard to structural invariance, traditional Chi^2 difference test approach was performed, indicating the QCSS's factorial invariance across gender ($\Delta\chi^2$ = 11.97, Δdf = 8, p = 0.152).

Self-Concepts and Attitudes Towards School

An adapted French version of the *CoSoi* (Valls and Bonvin, 2021) was used to measure self-concepts and attitudes towards school. Due to the age of the students involved in this study, the pictures were adapted with only one statement per item. This self-report questionnaire is composed of 13 items divided into four subscales: self-concept in language learning (SC-L; 3 items, α = 0.64) and in mathematics (SC-M; 3 items, α = 0.77), social self-concept (SC-Social; 3 items, α = 0.75) and Attitudes towards school (Attitudes; 3 items, α = 0.85). Academic self-concept corresponds to the student's evaluation of his/her general academic competence (i.e., in the two subjects mentioned). Social self-concept corresponds to the student's evaluation of his/her social relationships within the classroom, while attitudes towards school correspond to the student's evaluation of his/her emotional well-being at school. Each item was scored on a 4-point Likert scale with six reverse scored items (1 = "Not at all like me" to 4 = "Totally like me"; scores per subscale range from 3 to 9) and higher scores indicate higher self-concept (i.e., language learning, mathematics or social) and positive attitudes towards school. CFA results showed a good fit to the data

[1]https://www.cer-vd.ch/

[2]https://www.fedlex.admin.ch/eli/cc/2013/617/en

[3]https://prestations.vd.ch/pub/blv-publication/actes/consolide/172.65?key=1543934892528&id=cf9df545-13f7-4106-a95b-9b3ab8fa8b01

[4]https://etudiant.hepl.ch/files/live/sites/files-site/files/filiere-ps/programme-formation/code-ethique-recherche-cdhep-2002-fps-hep-vaud.pdf

[5]https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf

TABLE 1 | Means (M), standard deviations (SD) and *t*-test results for gender differences.

	Boys (*n* = 124)		Girls (*n* = 114)		Means comparison			
	M	**SD**	**M**	***SD***	***t***	***df***	***P***	***d***
SC-L	2.89	0.65	3.06	0.62	–2.08	236	0.039	0.27
SC-M	3.36	0.66	3.12	0.79	2.59	220.49	0.010	0.33
SC-Social	3.37	0.78	3.45	0.73	–0.80	236	0.423	—
Attitudes	2.74	0.99	3.19	0.86	–3.75	234.93	0.000	0.55
Upward contrast	1.84	0.75	2.21	0.90	–3.39	219.96	0.001	0.46
Upward identification	2.82	0.75	2.75	0.76	0.67	236	0.505	—
Downward identification	1.65	0.74	1.89	0.85	–2.32	236	0.021	0.30
Downward contrast	2.11	0.91	2.20	0.88	–0.82	236	0.412	—

Note. SC-L, language learning self-concept; SC-M, mathematics self-concept; SC-S, social self-concept; Attitudes, attitudes towards school.

TABLE 2 | Correlations between the variables of interest for girls (below the diagonal) and boys (above the diagonal).

	1.	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**
1. SC-L	—	0.08	0.22*	0.12	–0.32***	–0.15	–0.36***	0.05
2. SC-M	0.13	—	0.21*	0.29**	–0.21*	–0.11	–0.19*	–0.10
3. SC-Social	0.19*	0.07	—	0.13	–0.34***	–0.02	–0.32***	–0.08
4. Attitudes	0.21*	0.32***	0.16	—	–0.29**	0.05	–0.08	–0.16
5. Upward contrast	–0.34***	–0.26**	–0.35***	–0.12	—	0.28**	0.46***	0.30***
6. Upward identification	–0.15	0.01	–0.14	0.07	0.22*	—	0.25**	0.13
7. Downward identification	–0.23*	–0.28**	–0.36***	–0.15	0.49***	0.32***	—	0.08
8. Downward contrast	–0.09	–0.02	–0.23*	–0.18	0.48***	0.36***	0.25**	—

Note. SC-L, language learning self-concept; SC-M, mathematics self-concept; SC-S, social self-concept; Attitudes, attitudes towards school.
*p < *0.05.*
**p < *0.01.*
***p < *0.001.*

($\chi^2/df = 1.99$; CFI = 0.95; RMSEA = 0.07; SRMR = 0.06) with factorial invariance across gender ($\Delta\chi^2 = 4.99$, $\Delta df = 8$, $p = 0.758$).

Items of the two questionnaires were presented alternately (i.e., one item from the CoSoi, then one item from the QCSS and so on), making the overall questionnaire contained 25 items. This was done to prevent students from trying to be consistent in their responses. Respondents were asked to indicate how similar they thought they were to the student described in each statement. The general instruction stated that there were no right or wrong answers.

Statistical Analysis

In order to explore gender differences, Student's *t*-test was conducted using Cohen's *d* to assess effect sizes (Cohen, 1988). A stepwise linear regression analysis was carried out in order to determine the predictors of each dimension of academic self-concept. Model 1 of PROCESS (Hayes, 2013) was used to test the moderating effect of gender in the relationship between significant predictors and each dimension of academic self-concept. Then, stepwise multiple linear regression analyses were performed separately for girls and boys to assess the influence of SC-Social, Attitudes, and four social comparison processes on each dimension of academic self-concept (i.e., SC-L and SC-M). Variance Inflation Factor (VIF) values were examined, with a VIF value equal to or greater than 10.00 indicating a multicollinearity problem (Chatterjee et al., 2000). The VIF values were all less than 2.00 in all models tested.

RESULTS

Table 1 shows the means and standard deviations of study's variables for boys and girls. At the descriptive level, we can see that the preferred social comparison processes differ somewhat by gender. Although they both report first using upward identification, girls report using upward and downward contrast equally, whereas boys report using more downward contrast and then upward contrast. The process least reported by both girls and boys is downward identification. Results of *t*-test (**Table 1**) indicate that girls have better SC-L and more positive attitudes towards school, while boys have better SC-M (with small to moderate effect sizes). In addition, girls report using upward contrast and downward identification significantly more frequently than boys (with small effect sizes).

A first step was to explore correlations according to gender, which are reported in **Table 2**. It appeared that the strength of the associations between social comparison processes and dimensions of academic self-concept did not vary notably by gender. The only differences found were in the relationship between Attitudes and SC-L (the correlation being significant for girls but not for boys) and between SC-S and SC-M (the correlation being significant for boys but not for girls). Results of the first stepwise multiple linear regression analysis showed that upward contrast ($\beta = -0.25$, $p < 0.001$, $sr^2 = -0.21$), gender ($\beta = 0.21$, $p < 0.001$, $sr^2 = 0.21$) and downward identification ($\beta = -0.18$, $p < 0.001$, $sr^2 = -0.16$) emerged as significant predictors of

TABLE 3 | Hierarchical multiple linear regression analyses predicting self-concept in language learning (SC-L) across gender.

		R^2	Predictors	β	*t*	*p*	*F* change (*df*)
Boys	Step 1	0.06	SC-Social	0.20	2.30	0.023	3.63 (2,121)
			Attitudes	0.10	1.10	0.274	
	Step 2	0.19	SC-Social	0.08	0.86	0.394	4.69 (4,117)
			Attitudes	0.07	0.77	0.444	
			Upward contrast	–0.19	–1.75	0.083	
			Upward identification	–0.05	–0.60	0.551	
			Downward identification	–0.24	–2.50	0.014	
			Downward contrast	0.15	1.69	0.094	
Girls	Step 1	0.07	SC-Social	0.16	1.76	0.081	4.14 (2,111)
			Attitudes	0.18	1.96	0.052	
	Step 2	0.18	SC-Social	0.06	0.59	0.557	3.45 (4,107)
			Attitudes	0.20	2.16	0.033	
			Upward contrast	–0.35	–3.06	0.003	
			Upward identification	–0.13	–1.34	0.182	
			Downward identification	–0.01	–0.12	0.902	
			Downward contrast	0.17	1.61	0.111	

Note. SC-L, language learning self-concept; SC-M, mathematics self-concept; SC-S, social self-concept; Attitudes, attitudes towards school.

TABLE 4 | Hierarchical multiple linear regression analyses predicting self-concept in mathematics (SC-M) across gender.

		R^2	Predictors	β	*t*	*p*	*F* change (*df*)
Boys	Step 1	0.11	SC-Social	0.17	2.01	0.046	7.78 (2,121)
			Attitudes	0.27	3.10	0.002	
	Step 2	0.14	SC-Social	0.14	1.44	0.153	0.83 (4,117)
			Attitudes	0.26	2.88	0.005	
			Upward contrast	–0.02	–0.15	0.882	
			Upward identification	–0.09	–1.00	0.319	
			Downward identification	–0.10	–0.96	0.338	
			Downward contrast	–0.02	–0.21	0.832	
Girls	Step 1	0.10	SC-Social	0.01	0.15	0.880	6.13 (2,111)
			Attitudes	0.31	3.43	0.001	
	Step 2	0.21	SC-Social	–0.10	–1.04	0.302	3.53 (4,107)
			Attitudes	0.30	3.34	0.001	
			Upward contrast	–0.25	–2.26	0.026	
			Upward identification	0.03	0.33	0.745	
			Downward identification	–0.20	–1.88	0.063	
			Downward contrast	0.17	1.58	0.117	

Note. SC-L, language learning self-concept; SC-M, mathematics self-concept; SC-S, social self-concept; Attitudes, attitudes towards school.

SC-L ($F_{(3,234)}$ = 13.43, $p < 0.001$, R^2 = 0.15). However, the interaction effects of upward contrast and gender as well as downward identification and gender were not significant (p = 0.950 and p = 0.139, respectively). A second analysis was performed to predict SC-M ($F_{(3,234)}$ = 14.46, $p < 0.001$, R^2 = 0.16), and results showed that Attitudes ($\beta = 0.28$, $p < 0.001$, sr^2 = 0.27), downward identification (β = −0.21, $p < 0.0501$, sr^2 = −0.21) and gender (β = −0.20, $p < 0.01$, sr^2 = −0.19) were significant predictors. Results also showed that interaction effect of Attitudes and gender was not significant (p = 0.292), nor was the interaction effect of downward identification and gender (p = 0.460).

Results of stepwise multiple linear regression analyses are presented in **Tables 3**, **4**. As show in **Table 3**, only downward identification (sr^2 = −0.21) was a significant and negative predictor for boys regarding the prediction of SC-L, indicating that the more they would tend to use this process the lower their SC-L would be. For girls, beyond the positive contribution of Attitudes (sr^2 = 0.19), a negative effect of upward contrast (sr^2 = −0.27) on SC-L is found. Concerning the prediction of SC-M (see **Table 4**), a negative effect of upward contrast (sr^2 = −0.20) is found beyond the positive contribution of Attitudes (sr^2 = 0.29) for girls, while for boys only Attitudes are a significant predictor (sr^2 = 0.25).

DISCUSSION

This study had two main purposes: 1) to explore gender differences in self-concept (academic subject-specific and non-academic), attitudes towards school and social comparison processes; and 2) to explore the relative contribution of social comparison processes to academic subject-specific self-concepts by gender, beyond the influence of social self-concept and

attitudes towards school. Our results confirmed our hypotheses that girls in our sample engaged in more negative social comparison processes than boys (i.e., upward contrast and downward identification) and that upward identification was not a significant predictor of academic self-concept (for any of the observed subjects).

Regarding gender comparisons, results on academic self-concept are not surprising given that several studies have shown higher levels of mathematics self-concept among boys while girls reported higher levels of language learning self-concept (e.g., Bouffard et al., 2006; Marsh, 1989). Some studies, conducted with culturally similar samples to the present study, had nevertheless found that the social comparison processes preferentially used by students were upward identification and downward contrast (Boissicat et al., 2012; Bouffard et al., 2014). However, our results indicate that while we do find these preferences for boys, girls use upward and downward contrast equally after upward identification. As for downward identification, it is the least used process, regardless of gender.

Results of stepwise multiple linear regression analyses show that upward contrast best explains gender differences, with a stronger effect for girls. Attitudes towards school only explain gender differences in language learning self-concept. Furthermore, positive processes (i.e., upward identification and downward contrast) have no effect on either component of academic self-concept. Yet, Boissicat et al. (2012) reported a stronger negative influence of downward identification while upward identification had a positive but small contribution. These differences in results can be explained on the one hand by the fact that we conducted separate analyses by gender. Furthermore, it is clear that upward contrast (i.e., contrast with another student judged as having superior abilities) has a deleterious effect on self-concept in mathematics and language learning for girls, while it is only marginally significant for boys in language learning ($p = 0.08$). Thus, as suggested by Dumas and Huguet (2011), it would appear that the supposed positive effects of upward identification are insufficient to counteract the negative effects of upward contrast especially for girls and regardless of the subject concerned. The non-significant moderating effects of gender indicate that the effects of social comparison processes on subject-specific self-concepts do not vary substantially between girls and boys, as suggested by the correlations according to gender. The gender differences found in the social comparison processes could be explained by other moderating variables not taken into account in this study, such as "social comparison orientation" (SCO; Gibbons and Buunk 1999), which has not yet been studied in elementary school students (Dijkstra et al., 2008). SCO refers to "the extent to which and the frequency with which people compare themselves with others" (Dijkstra et al., 2010, p. 196). It turns out that individuals with a high SCO would seek out more social comparison and that these processes would also affect them more negatively (Buunk and Gibbons, 2006). Thus, it is possible that girls may be more sensitive to SCO. Furthermore, Bouffard et al. (2014) report moderate correlations (i.e., about 0.40) between negative social comparison processes (i.e., upward contrast and downward identification) and school anxiety but unfortunately, they did not compare levels of such anxiety according to gender. We can nevertheless suppose that girls may have higher levels of SCO than boys, with negative social comparison processes having a greater impact on them and a greater risk of experiencing school anxiety in relation to the subject influenced by gender stereotypes (e.g., mathematics). Moreover, for girls, although upward contrast contributes to explain most of the variance in mathematics self-concept, downward identification is marginally significant ($p = 0.06$) while neither is significant for boys.

Limitations of the Study

However, the major limitation of our study is the induction of forced social comparison. Indeed, the latter can only be hypothetical and can therefore differ greatly from the voluntary and deliberate comparison carried out within the classroom, that is in which students can choose a real classmate with whom they compare themselves (Boissicat et al., 2020). In addition, forced social comparison may lead students to compare themselves on a dimension that is of little or no relevance to them, with the results producing effects not comparable to those obtained with deliberate social comparisons in subjects perceived as relevant by students (Dijkstra et al., 2008). Nevertheless, both types of social comparison (i.e., forced and deliberate) are important to consider because they can coexist in the classroom context. Generally, during forced comparisons, upward contrast effects are predominant, but upward identification effects may be added during deliberate comparisons for adaptive purposes (Dumas and Huguet, 2011). In particular, a qualitative study of 246 students between the ages of 10 and 11 found that forced comparisons were less common in the classroom setting, with students reporting that they were more likely to compare themselves if they could choose a friend to do so with (Webb-Williams, 2021). Moreover, forced comparison would only take place if they are struggling. Thus, when they deliberately compare themselves, students would choose targets of the same sex with a tendency to compare upward (Dumas and Huguet, 2011; Boissicat et al., 2020). The effects of upward social comparison are therefore complex in nature, and also depend on the type of comparison (forced vs. deliberate). According to Dumas and Huguet (2011), when students actively seek it out, its influence would be beneficial for academic self-concept and academic achievement. Moreover, not having considered students' academic achievement and grade-point average of classrooms is another limitation. For example, Webb-Williams (2021) found that students in low-ability group were more vulnerable to the negative effects of social comparison on self-evaluation and performance evaluation, and avoided upward comparisons. Knowing that the effect of social comparison on academic self-concept may be due to the Big-Fish-Little-Pond Effect (BFLPE; Marsh and Parker 1984), it is possible that high-achieving students may self-assess their competence as average or inferior if they are in a high-achieving class or school (and inversely). Yet, Huguet et al. (2009) showed that BFLPE was

rooted in the disadvantageous way students compared themselves to most of their classmates (i.e., forced under the pressure of the environment), but that beyond these comparisons students made comparison choices (i.e., deliberate and for adaptive purposes) that sometimes had a beneficial effect on their academic self-concept. Returning to the previously mentioned findings of Webb-Williams (2021), grouping students by ability levels would imply a forced comparison with students of the same levels, restricting the potential positive effect of deliberate comparison. All of these elements may thus explain why the positive effects of social comparison processes did not emerge in the results of the present study. A final limitation is the cross-sectional design of the study. Although Wolff et al. (2018) showed that social comparison had a stronger effect on academic self-concept than dimensional or temporal comparison, it seems important to conduct longitudinal studies that include multiple measurement times over a school year and within the same classroom to observe variations in social comparison processes and SCO levels across subjects.

CONCLUSION

Thus, recall the conclusion of the meta-analysis by Gerber et al. (2018), namely: "The common response to comparison is contrast: people increase their self-evaluations after downward comparison and decrease their self-evaluations after upward comparisons." (p. 194). Even though our results must be interpreted with caution, it appears that upward contrast best explains gender differences in our sample, and its negative effect is not reduced by less frequent use or by more frequent implementation of others processes. Results of this study demonstrate the need to examine the evolution of social comparison processes over time, considering their impact on achievement as well as on students' academic and social well-being from a gender perspective. This would also allow us to explore the existence of particular profiles and to assess their risk in order to implement strategies to limit their negative impacts on students. It also seems necessary to take into consideration different motivations for comparison (e.g., self-assessment, improvement, valorization) and its level of orientation (i.e., SCO). Making teachers aware of the existence of these social comparison processes appears to be an important issue. Knowing that some teachers may deliberately use social comparison within their classrooms (i.e., forced comparison), having knowledge about the influence of these processes on academic self-concept also seems relevant, as already highlighted by Wolff et al. (2018). These authors also indicate that comparison should above all be redirected to internal and temporal frames of reference with positive reinforcements and by highlighting opportunities for improvement among below-average performing students in order to strengthen their academic self-concept. This may also help them to objectively self-assess their academic competence.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not provided by the participants' legal guardians/next of kin because Parents were informed by letters of the general objectives of the study, and could decline their child's participation in the data collection. Under this condition, no refusals were recorded and the anonymity of the participants was preserved.

AUTHOR CONTRIBUTIONS

MV is the sole author of this submission and is accountable for the content of this work.

ACKNOWLEDGMENTS

Particular acknowledgments is given to the Bachelor students for their support in collecting data.

REFERENCES

Boissicat, N., Pansu, P., Bouffard, T., and Cottin, F. (2012). Relation between Perceived Scholastic Competence and Social Comparison Mechanisms Among Elementary School Children. *Soc. Psychol. Educ.* 15, 603–614. doi:10.1007/s11218-012-9189-z

Boissicat, N., Pansu, P., and Bouffard, T. (2020). Does Classroom Social Comparison Bias Students' Evaluation of Their Own Competence? *Soc. Psychol. Educ.* 23, 1303–1326. doi:10.1007/s11218-020-09582-y

Bouffard, T., Pansu, P., Boissicat, N., Vezeau, C., and Cottin, F. (2014). Questionnaire de la comparaison de soi scolaire pour enfants et adolescents. *Can. J. Behav. Sci./Revue canadienne des Sci. du comportement* 46, 84–94. doi:10.1037/a0028429

Bouffard, T., Vezeau, C., Chouinard, R., and Marcotte, G. (2006). L'illusion d'incompétence et les facteurs associés chez l'élève du primaire*. *rfp* 155, 9–20. doi:10.4000/rfp.61

Buunk, A. P., and Gibbons, F. X. (2006). "Social Comparison Orientation: A New Perspective on Those Who Do and Those," in *Social Comparison and Social Psychology. Understanding Cognition, Intergroup Relations, and Culture*. Editor S. Guimond (New York: Cambridge University Press), 15–32.

Buunk, B. P., and Gibbons, F. X. (2000). "Toward an Enlightenment in Social Comparison Theory," in *Handbook of Social Comparison: Theory and Research*. Editors J. Suls and L. Wheeler (Boston: Springer), 487–499. doi:10.1007/978-1-4615-4237-7_22

Buunk, B. P., Kuyper, H., and Van der Zee, Y. G. (2005). Affective Response to Social Comparison in the Classroom. *Basic Appl. Soc. Psychol.* 27, 229–237. doi:10.1207/s15324834basp2703_4

Chatterjee, S., Hadi, A. S., and Price, B. (2000). *Regression Analysis by Example*. New York: John Wiley & Sons.

Cohen, J. (1988). *Statistical Power Analysis in the Behavioral Sciences*. Hillsdale: Erlbaum.

Dijkstra, P., Gibbons, F. X., and Buunk, A. P. (2010). "Social Comparison Theory," in *Social Psychological Foundations of Clinical Psychology*. Editors J. E. Maddux and J. P. Tangney (New York: The Guilford Press), 195–211.

Dijkstra, P., Kuyper, H., Van der Werf, G., Buunk, A. P., and van der Zee, Y. G. (2008). Social Comparison in the Classroom: A Review. *Rev. Educ. Res.* 78, 828–879. doi:10.3102/0034654308321210

Dumas, F., and Huguet, P. (2011). "Chapitre 9 : Le double visage de la comparaison sociale à l'école," in *L'évaluation, Une Menace*. Editors F. Butera, C. Buchs, and C. Darnon (Paris: Presse Universitaires de France), 95–104. doi:10.3917/puf.darno.2011.01.0095

Ertl, B., Luttenberger, S., and Paechter, M. (2017). The Impact of Gender Stereotypes on the Self-Concept of Female Students in STEM Subjects with an Under-representation of Females. *Front. Psychol.* 8, 703. doi:10.3389/fpsyg.2017.00703

Gerber, J. P., Wheeler, L., and Suls, J. (2018). A Social Comparison Theory Meta-Analysis 60+ Years on. *Psychol. Bull.* 144, 177–197. doi:10.1037/bul0000127

Gibbons, F. X., and Buunk, B. P. (1999). Individual Differences in Social Comparison: Development of a Scale of Social Comparison Orientation. *J. Pers. Soc. Psychol.* 76, 129–142. doi:10.1037/0022-3514.76.1.12910.1037//0022-3514.76.1.129

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York: Guilford Press.

Huguet, P., Dumas, F., Marsh, H., Wheeler, L., Seaton, M., Nezlek, J., et al. (2009). Clarifying the Role of Social Comparison in the Big-Fish-Little-Pond Effect (BFLPE): an Integrative Study. *J. Pers Soc. Psychol.* 97, 156–170. doi:10.1037/a0015558

Marsh, H. W., and Parker, J. W. (1984). Determinants of Student Self-Concept: Is it Better to Be a Relatively Large Fish in a Small Pond Even if You Don't Learn to Swim as Well? *J. Personal. Soc. Psychol.* 47, 213–231. doi:10.1037/0022-3514.47.1.213

Marsh, H. W. (1989). Sex Differences in the Development of Verbal and Mathematics Constructs: The High School and beyond Study. *Am. Educ. Res. J.* 26, 191–225. doi:10.3102/00028312026002191

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/external Frame of Reference Model. *Am. Educ. Res. J.* 23, 129–149. doi:10.3102/00028312023001129

Pulford, B. D., Woodward, B., and Taylor, E. (2018). Do social Comparisons in Academic Settings Relate to Gender and Academic Self-Confidence? *Soc. Psychol. Educ.* 21, 677–690. doi:10.1007/s11218-018-9434-1

Rost, D. H., Sparfeldt, J. R., Dickhäuser, O., and Schilling, S. R. (2005). Dimensional Comparisons in Subject-specific Academic Self-Concepts and Achievements: a Quasi-Experimental Approach. *Learn. InstructionInstr* 15, 557–570. doi:10.1016/j.learninstruc.2005.08.003

Smith, R. H. (2000). "Assimilative and Contrastive Emotional Reactions to Upward and Downward Social Comparisons," in *Handbook of Social Comparison: Theory and Research*. Editors J. Suls and L. Wheeler (Boston: Springer), 173–200. doi:10.1007/978-1-4615-4237-7_10

Valls, M., and Bonvin, P. (2021). Évaluer le concept de soi des élèves en contexte scolaire inclusif : exploration des qualités psychométriques du CoSoi. *Neuropsychiatrie de l'Enfance et de l'Adolescence* 69, 176–184. doi:10.1016/j.neurenf.2021.04.002

Webb-Williams, J. L. (2021). Teachers' Use of Within-Class Ability Groups in the Primary Classroom: A Mixed Methods Study of Social Comparison. *Front. Psychol.* 12, 728104. doi:10.3389/fpsyg.2021.728104

Wolff, F., Helm, F., Zimmermann, F., Nagy, G., and Möller, J. (2018). On the Effects of Social, Temporal, and Dimensional Comparisons on Academic Self-Concept. *J. Educ. Psychol.* 110, 1005–1025. doi:10.1037/edu0000248

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Education

ORIGINAL RESEARCH
published: 17 January 2022
doi: 10.3389/feduc.2021.793486

Gender Stereotypes and Expected Backlash for Female STEM Students in Germany and Japan

Laura Froehlich[1], Saori Tsukamoto[2], Yasuko Morinaga[3], Kiriko Sakata[3], Yukiko Uchida[4], Melanie M. Keller[5], Stefan Stürmer[1], Sarah E. Martiny[6] and Gisela Trommsdorff[7]*

[1]Research Cluster D²L², FernUniversität in Hagen, Hagen, Germany, [2]Division of Liberal Arts and Sciences, Aichi Gakuin University, Nisshin, Japan, [3]Graduate School of Humanities and Social Sciences, University of Hiroshima, Hiroshima, Japan, [4]Kokoro Research Center, Kyoto University, Kyoto, Japan, [5]IPN - Leibniz Institute for Science and Mathematics Education, Kiel, Germany, [6]Department of Psychology, UiT the Arctic University of Norway, Tromsø, Norway, [7]Department of Psychology, University of Konstanz, Konstanz, Germany

Although Germany and Japan are top-ranking in STEM, women are underrepresented in the STEM fields of physics, engineering, and computer science in both countries. The current research investigated widespread gender-science stereotypes in STEM in the two countries (Studies 1 and 2) and negative consequences of expected backlash (i.e., imagining negative reactions and lower ascribed communion in scenarios) for women's emotions and motivation in STEM due to role incongruity and lack-of-fit (Study 3). Studies 1 (N = 87) and 2 (N = 22,556) showed that explicit and implicit gender-science stereotypes are widespread and comparable in Germany and Japan. Study 3 (N = 628) showed that lower ascribed communion was related to less positive emotions, more negative emotions and anxiety emotions, and less study motivation for STEM students (from the fields of physics, engineering, and computer science) from Germany and Japan. Results point to more subtle expected backlash effects for women in STEM than hypothesized. Theoretical and practical implications for gender equality in STEM are discussed.

Keywords: backlash, cross-cultural psychology, gender stereotypes, social role theory, science technology engineering mathematics

OPEN ACCESS

Edited by:
Jan Retelsdorf,
University of Hamburg, Germany

Reviewed by:
Angelica Moè,
University of Padua, Italy
Kathryn Everhart Chaffee,
Université du Québec à Montréal, Canada

****Correspondence:***
Laura Froehlich
laura.froehlich@fernuni-hagen.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *12 October 2021*
Accepted: *07 December 2021*
Published: *17 January 2022*

Citation:
Froehlich L, Tsukamoto S, Morinaga Y, Sakata K, Uchida Y, Keller MM, Stürmer S, Martiny SE and Trommsdorff G (2022) Gender Stereotypes and Expected Backlash for Female STEM Students in Germany and Japan. Front. Educ. 6:793486. doi: 10.3389/feduc.2021.793486

INTRODUCTION

Around the world, women are underrepresented in Science, Technology, Engineering, and Mathematics (STEM) fields. Across the member states of the Organization of Economic Co-operation and Development (OECD), 72% of engineering and 80% of information technology degrees are awarded to men (OECD, 2015). However, gender distributions differ between STEM fields. Whereas women's representation in biology, chemistry and mathematics is equal or even higher than men's, women are clearly underrepresented in physics, engineering, and computer science (e.g., Cheryan et al., 2017). Women's underrepresentation in these fields is unlikely to be explained by gender differences in mathematical ability, as numerous studies found that men and women show equal math performance (e.g., Else-Quest et al., 2010; Lindberg et al., 2010). The topic of gender differences in STEM has been investigated in numerous disciplines. Social-psychological research highlights how gender stereotypes and their consequences for women's emotion, motivation, and behavior contribute to their underrepresentation in STEM (e.g., Eagly and Karau, 2002; Eagly and Wood, 2012).

Whereas a large amount of social-psychological work on women's underrepresentation in STEM has focused on the United States (e.g., Cheryan et al., 2017; Diekman et al., 2017), the gender gap in STEM varies around the world. It is of increasing importance to investigate factors that contribute to cross-cultural differences and similarities in women's underrepresentation in STEM (e.g., Yalcinkaya and Adams, 2020). Therefore, the current research focuses on Germany and Japan, two top-ranking countries in STEM, which for example, are among the top 5 countries in natural-science research (Nature Index, 2020) and technological expertise (U. S. News and World Report LP, 2020). Despite their success in STEM, in both countries women are underrepresented in physics, engineering, and computer science. In these fields, less than one third of undergraduate students were female (Germany: physics: 30%, engineering: 24%, computer science: 21%; Japan: science: 27%, engineering: 14%; Destatis, 2019; Gender Equality Bureau Cabinet Office, 2017).

Social psychological research has shown that gender stereotypes associate STEM with males (e.g., Nosek et al., 2009) and that women entering counter-stereotypic fields can experience social repercussions in form of backlash effects (e.g., Rudman and Glick, 2001). The current research investigates how gender stereotypes and expected backlash effects contribute to the gender gap in STEM in Germany and Japan, two different cultural contexts in which group membership is of varying relevance to the individuals (Markus and Kitayama, 1991) and which have received less scholarly attention than the cultural context of the United States.

Gender Stereotypes in STEM

Despite gender similarities in performance, women's STEM abilities and motivation are stereotyped as low in many countries (e.g., Miller et al., 2015; Nosek et al., 2009). Stereotypes are "beliefs and associations that link a whole group of people with certain traits or characteristics" (Kassin et al., 2011, p. 148) and can be described on the dimensions of agency and communion (e.g., Williams and Best, 1991). Agency consists of competence ("capable") and assertiveness ("ambitious"), whereas communion consists of warmth ("friendly") and morality ("honest"; Abele et al., 2016). Men are stereotyped as agentic and women as communal (e.g., Williams and Best, 1991). As STEM is stereotypically associated with traits that are more valued in men than in women (Cheryan et al., 2015), negative stereotypes about women's agency likely have detrimental consequences for women in STEM. They are associated with lower domain identification, career intentions (e.g., Cundiff et al., 2013), interest, sense of belonging (e.g., Cheryan et al., 2009), and lower enrollment in STEM classes (e.g., Stout et al., 2016). Thus, it can be assumed that gender stereotypes contribute to women's underrepresentation in STEM.

Research conducted separately in Germany and Japan showed that women are negatively stereotyped in STEM in both countries (e.g., Adachi, 2014; Ikkatai et al., 2020; Steffens and Jelenec, 2011). The current research conducts a joint investigation of gender stereotypes in these two countries to gain knowledge about potential similarities and differences in gender stereotypes and their psychological consequences for (female) STEM students. Further, we aim to study whether the psychological processes that are related to widespread gender stereotypes and women's underrepresentation in STEM are generalizable in these two countries representing different world regions: Whereas Germany can be categorized as a WEIRD (i.e., Western, Educated, Industrialized, Rich, Democratic), the East Asian country of Japan, although rich and industrialized, is commonly classified as non-Western (Henrich et al., 2010).

A recent model of cross-national variation in gender gaps in STEM participation (Yalcinkaya and Adams, 2020) proposed that individualistic, post-materialistic WEIRD countries show higher underrepresentation of women in STEM than collectivistic, materialistic countries. The model explains national differences in STEM gender gaps by differences in values emphasizing individual choice vs. financial security and relational expectations. However, there are deviations from this proposed dichotomy, embodied by Germany and Japan. Germany is more individualistic than Japan (e.g., Varnum et al., 2010), but the structural and economic factors are similar: Both countries are industrialized and affluent (e.g., Credit Suisse Research Institute, 2019), and in both countries the gender gap in STEM is large. As gender stereotypes arise from the gendered division of labor (e.g., Eagly and Wood, 2012) and women are underrepresented in STEM in Germany and Japan, we thus expect STEM ability to be stereotypically associated with men rather than women in both countries (Hypothesis 1; Studies 1 and 2). Focusing on Germany and Japan, we aim at investigating which aspects of the consequences of gender stereotypes for women in STEM are generalizable across countries and whether they are related to cultural variables reflecting the relevance of social group membership and associated stereotypes for the self (Study 3).

Backlash and Lack of Fit for Women in STEM

Social-psychological theories describe negative consequences of gender stereotypes for women in male-dominated domains (e.g., leadership, STEM). Social role theory (e.g., Eagly and Wood, 2012) posits that gender stereotypes arise because men and women occupy different social roles. The observation of gender-segregated social roles leads to stereotypes, which subsequently influence motivation, emotion, and behavior. Higher role segregation and stronger stereotypes lead to gender differences in behavior. Women (men) are expected to behave communal (agentic). However, women pursuing a STEM career behave counter-stereotypically, which can lead to negative social consequences like being perceived as unlikable.

The lack-of-fit framework describes that social roles stereotyped to require agentic traits (e.g., leadership positions) are perceived as incongruent with the female stereotype, resulting in a perceived lack-of-fit of women with these roles (e.g., Heilman, 1983). According to role congruity theory (Eagly and Karau, 2002) men's roles, but not women's, overlap with leadership roles. When women enter a field stereotyped as agentic or display agentic behavior–thereby violating

prescriptive gender stereotypes (i.e., how women should behave; e.g., Eagly and Karau, 2002), they likely experience a backlash effect (i.e., social repercussions for counter-stereotypical behavior). Agentic women receive negative social reactions in that they are evaluated as socially deficient and unlikable (low in communion) by others (Rudman and Phelan, 2008).

Based on social role theory and role congruity theory, we investigate how women in STEM expect backlash as a consequence of gender stereotypes. STEM fields, especially physics, engineering and computer science, are stereotypically associated with men (e.g., Cheryan et al., 2015) and work in STEM fields is not perceived as people-oriented (e.g., Gino et al., 2015), representing communal work goals (Cheryan et al., 2017). Therefore, we expect women in these STEM fields to expect backlash (Hypothesis 2; Study 3). The current research focuses on expected rather than experienced backlash for several reasons. First, investigating actual backlash behavior (repercussions for counter-stereotypical behavior from other people) would require an observational or experimental methodology, which was beyond the scope of the survey conducted in Study 3. Second, the focus of the current research was on how female students expect backlash due to their study major and how this subjective perception of potential backlash influences their subsequent emotions and motivation. We believe that this focus on subjective expectations of backlash is highly relevant, as these subjective expectations are likely to be a proximal predictor of emotions and motivation.

Negative social reactions can in turn influence women's emotions and motivation in STEM. Morinaga et al. (2017) investigated how benevolent sexism affects women's emotions and motivation in mathematics in two scenario experiments with Japanese female (junior) high-school students. When students imagined their math teacher to comment a good performance with "well done, although you are a girl!" (stereotype activation condition), they experienced more negative and less positive emotions than in a control condition ("well done!"). Stereotype activation lead to lower motivation mediated by emotions. In line with this, we expect that for women, but not for men, the expectation of more negative reactions to studying a STEM subject (expected backlash effect) is related to negative emotions (Hypothesis 3). In turn, these emotions predict lower motivation to study for STEM (Hypothesis 4).

A Cross-Cultural Approach to Expected Backlash for Female STEM Students

If STEM is stereotypically associated with men in both Germany and Japan, it is likely that these stereotypes have negative psychological consequences for female STEM students in both cultural contexts. The negative consequences of backlash effects have been predominantly investigated in the United States (e.g., Rudman and Glick, 2001; Rudman and Phelan, 2008; Eaton et al., 2020). It remains unclear whether the expected negative reactions for counter-stereotypical behavior are related to women's emotion and motivation in a similar way and intensity in cultural contexts in which membership in social groups is of varying relevance to the self. To fill this gap in the literature, the present research investigates gender stereotypes and their psychological consequences for women in stereotype-incongruent STEM fields in Germany and Japan and examines whether the psychological variable of self-construal, which reflects how central social group membership is for the self, is associated with the extent of expected backlash effects.

In the Japanese culture individuals tend to endorse an interdependent self-construal, a cultural orientation for which social group membership is central to the self. In the German culture individuals tend to endorse an independent self-construal, for which group membership is less central (Markus and Kitayama, 1991; Varnum et al., 2010). In cultures where individuals tend to endorse an interdependent self-construal, social networks are relatively stable (i.e., low relational mobility; Thomson et al., 2018) and people are highly sensitive to social rejection (e.g., Sato et al., 2014). We thus argue that self-construal is relevant when investigating expected backlash effects of female STEM students across cultures, as individuals endorsing an interdependent self-construal should be more prone to expecting negative social repercussions for their counter-stereotypical behavior (studying a STEM subject) than individuals endorsing an independent self-construal. We thus expect that the kind of self-construal moderates the effects of expected backlash on female STEM students' emotions and motivation (Hypothesis 5). We explore whether these relationships depend on the relational mobility afforded by the social situation. Associations between variables should be stronger in a low relational mobility situation (new relationships are likely to become stable) compared to a high relational mobility situation (relationships are flexible and formed by personal choice).

The Present Research

As a basis for the investigation of the consequences of gender stereotypes for female STEM students in Germany and Japan, in a first step (Studies 1 and 2) we aim at substantiating that in both countries gender stereotypes associate men more with STEM than women. Study 1 investigates explicit gender stereotypes about mathematical and general academic abilities. Because explicit measurement of stereotypes can be prone to response biases (e.g., Smith, 2014; Kemmelmeier, 2016), Study 2 investigates explicit and implicit gender-science stereotypes using samples from Project Implicit. In a second step, we investigate expected backlash effects for female students of physics, engineering and computer science for the first time jointly in Germany and Japan. Study 3 (pre-registered) investigates the consequences of gender stereotypes for German and Japanese STEM students. In two scenarios, participants were asked to imagine a conversation with a previously unknown person of the opposite gender who is asking about their field of study. The participants indicated how they expected their conversation partner to react and perceive them on communion. We hypothesize that women expect more negative reactions and lower communion ratings than men (expected backlash). Furthermore, expected backlash should have negative consequences for women's emotions and motivation in STEM and should be stronger for individuals

strongly endorsing an interdependent self-construal. Materials (Studies 1 and 3), data and analysis scripts (all studies), and the pre-registration (Study 3) are available on the OSF (https://osf.io/4awqe/).

STUDY 1: EXPLICIT GENDER-MATH STEREOTYPES

To replicate the basic premise that men are more strongly associated with STEM and high STEM ability than women (e.g., Steffens and Jelenec, 2011; Ikkatai et al., 2019) in both countries, in a questionnaire study we assessed participants' perceptions of widespread gender stereotypes about math and general academic abilities.

Methods

Data were collected in December 2013 (Japan) and September 2015 (Germany). University students were recruited as participants via e-mail, a virtual laboratory and in class. Participants did not receive compensation for participation. The sample consisted of 28 Japanese (age: M = 26.15 years, SD = 7.34, 42.9% female) and 59 German university students (age: M = 33.25 years, SD = 10.18, 74.6% female). Participants answered a questionnaire assessing gender stereotypes about math and general academic abilities and their valence. Materials were translated and back-translated by the research team. Participants listed stereotypical statements about women's and men's general academic and math abilities and rated the statements' valence (from −3 = very negative to +3 = very positive). Participants were asked not to provide their personal opinion, but indicate socially shared stereotypes in Germany or Japan. Finally, they provided demographic information (age, gender, nationality) and were debriefed.

Results

Stereotype Content

Japanese participants made 221 statements (women/math: 61, women/general: 55, men/math: 55, men/general: 50), and German participants made 924 statements (women/math: 218, women/general: 239, men/math: 228, men/general: 239). In both samples, most statements about women's math ability indicated a negative conception. For example, participants indicated "slow in doing mental arithmetic," "bad at logical thinking/algebra." In contrast, for women's general academic ability, participants mostly indicated that they are good at languages and humanities, for example, "good at languages" or "good at arts and music." Men's math ability was described with positive statements, e.g., "good at math/logical thinking," "good comprehension of mathematical formulas." In turn, men's general academic ability was characterized as "good at math and natural science" or "bad at languages." The statements reflected the widespread stereotype that women have high abilities in languages and humanities but low abilities in math and science, and vice versa for men (e.g., Steffens and Jelenec, 2011).

Stereotype Valence

Valence ratings were averaged for each category. Ratings were nested within participants, we therefore computed a linear mixed model. To do so, we transformed the data from wide format (1 row per participant) to wide format (4 rows per participant, reflecting repeated measures of Domain and Gender). Because many participants listed less than the maximum number of five statements per category, we used restricted maximum likelihood (REML) estimation as it can produce unbiased estimates of variance and covariance parameters in the presence of missing data and uses the full data set; in contrast to full maximum likelihood estimation with listwise deletion. The dependent variable was valence ratings, predictors were Gender (male vs. female, within participants), Domain (general academic vs. math, within participants), and Country (Germany vs. Japan, between participants). Main and interaction effects were entered as fixed effects, the covariance type was compound symmetry. The main effect of Country was non-significant, F (1, 87.17) = 0.20, p = 0.657. There were significant main effects of Gender, F (1, 231.95) = 54.51, $p < 0.001$, and Domain, F (1, 253.04) = 5.49, p = 0.020. The interaction of Domain and Gender was also significant, F (1, 231.95) = 29.59, $p < 0.001$. The interactions with country were non-significant, Fs < 0.87, ps > 0.351. Bonferroni-adjusted post-hoc comparisons for the interaction of Gender and Domain across countries revealed that women's math ability was rated significantly more negatively than men's [M_{women} = −0.86, 95% CI (−1.16; −0.55), SE = 0.15, M_{men} = 1.04 (0.73; 1.34), SE = 0.15; t (223.63) = 9.37, SE = 0.20, $p < 0.001$, Cohen's d = 1.58]. Valence of women and men's general academic abilities did not differ significantly [M_{women} = 0.30, 95% CI (−0.01; 0.60), SE = 0.16; M_{men} = 0.58 (0.26; 0.91), SE = 0.16; t (239.34) = 1.33, SE = 0.22, p = 0.183, d = 0.23]. Across countries women's math ability was rated more negatively than their general academic ability [t (238.86) = 5.59, SE = 0.21, $p < 0.001$, d = 0.90]. Men's math ability was rated more positively than their general academic ability [t (246.25) = 2.12, SE = 0.21, p = 0.035, d = 0.41].

Discussion

In line with previous studies conducted separately in Germany and Japan (Ikkatai et al., 2019; Steffens and Jelenec, 2011), findings indicate the presence of negative stereotypes about women's math ability in Japanese and German society (Hypothesis 1). Participants indicated that women's math ability is stereotyped more negatively than men's, and also more negatively than women's general academic ability. These effects can be considered large (Cohen, 1988). There were no country differences between stereotype content and valence ratings. However, samples were small and stereotypes were measured only explicitly. To rule out response bias in explicit stereotype measurement (e.g., Smith, 2014; Kemmelmeier, 2016), in Study 2, we investigated gender-science stereotypes with data from Project Implicit.

STUDY 2: EXPLICIT AND IMPLICIT GENDER-SCIENCE STEREOTYPES

Study 2 investigated explicit and implicit gender-science stereotypes in Germany and Japan by Project Implicit (https://implicit.harvard.edu), which provides different Implicit Association Tests (IAT; Greenwald et al., 1998) to the public

in various languages. The gender-science IAT is a behavioral task measuring the implicit association between the categories male/female and science/liberal arts. Participants from 34 countries who completed gender-science IATs on the Project Implicit website associated male with science and female with liberal arts more easily than the reverse category combination (Nosek et al., 2009).

Method

Data provided by Project Implicit contained responses from 72,094 participants. Participants with missing values on the measure of implicit gender-science association (n = 44,010), missing values on gender (n = 4,017), or an age below 18 years (n = 1,159) were excluded. The final sample (N = 22,556) consisted of 9,875 Japanese (age: 18–88 years, M = 28.46, SD = 10.23; 50% female) and 12,681 German participants (age: 18–87 years, M = 29.54, SD = 10.22, 54% female).

Participants completed the gender-science IAT between 2006 and 2017. They categorized words into four categories by pressing two keys. In a stereotype-congruent condition the categories male/science were paired on one key and female/liberal arts on the other; in the stereotype-incongruent condition the pairings were reversed. Faster responses in the stereotype-congruent condition compared to the stereotype-incongruent condition indicate a stronger male-science association. Details on Project Implicit's gender-science IAT procedure can be found in Nosek et al. (2009). In addition, participants responded to the item "How much do you associate science with males or females" (1 = strongly male to 7 = strongly female) as a measure of explicit gender-science stereotypes, and provided demographics.

Statistical Analyses

In contrast to Study 1, which included a mixed model with between- and within-participants factors, Study 2 predicted implicit and explicit stereotypes by the between-participants factors Gender and Country. To do so, we used factorial Analysis of Variance (ANOVA). As only participants who completed the IAT were included in the sample, there were no missing values in the analysis of implicit stereotypes. For explicit stereotypes, a subsample of 51% of participants who completed the IAT also completed the explicit stereotype measure. Again, analyses were conducted with the subsample that completed the respective measure. In additional ANCOVAs, we controlled for year of data collection.

Results

Implicit Stereotypes

Project Implicit computed D scores as a measure of the implicit gender-science association for each participant by dividing the difference in mean response latency between the two conditions by the participant's latency standard deviation inclusive of the two conditions using the improved scoring algorithm (Nosek et al., 2009). Participants from Germany as well as from Japan showed positive overall D scores, indicating a stronger implicit association of male/science and female/liberal arts than the reverse combination [$M_{Germany}$ = 0.43, 95% CI (0.42; 0.44), SE = 0.01, M_{Japan} = 0.38 (0.37; 0.39), SE = 0.01]. We subjected the D scores to a 2 × 2 ANOVA with the between-participants factors Gender (men vs. women) and Country (Germany vs. Japan). Results showed significant main effects of Gender, F (1, 22,552) = 493.27, $p < 0.001$, Country, F (1, 22,552) = 70.49, $p < 0.001$, and a significant interaction, F (1, 22,552) = 145.89, $p < 0.001$. Bonferroni-adjusted post-hoc comparisons showed that in both countries, women showed stronger implicit associations of male/science and female/liberal arts than men [Germany: M_{Women} = 0.45 (0.44; 0.46), SE = 0.01, M_{Men} = 0.40 (0.39; 0.41), SE = 0.01, t (22,552) = 7.64, SE = 0.01, $p < 0.001$, d = 0.14; Japan: M_{Women} = 0.47 (0.46; 0.48), SE = 0.01, M_{Men} = 0.29 (0.27; 0.30), SE = 0.01, t (22,552) = 22.84, SE = 0.01, $p < 0.001$, d = 0.46]. For men, German participants showed a stronger implicit association than Japanese [t (22,552) = 14.25, SE = 0.01, $p < 0.001$, d = 0.14]. This difference was also significant for women, but with a small effect size [t (22,552) = 2.60, SE = 0.01, $p < 0.001$, d = 0.05]. Results were mainly robust when controlling for year of data collection (albeit the last comparison was no longer significant).

Explicit Stereotypes

The explicit gender-science stereotype item was completed by 11,601 participants (51% of the total sample). Means were above the scale midpoint, indicating that science was stereotyped to be male. A 2 × 2 ANOVA with Gender and Country as between-participants factors and explicit stereotypes as the dependent variable showed significant main effects of Gender, F (1, 11,597) = 89.02, $p < 0.001$, Country, F (1, 11,597) = 276.59, $p < 0.001$, and a significant interaction, F (1, 11,597) = 74.72, $p < 0.001$. Bonferroni-adjusted post-hoc comparisons showed that in Germany, men showed stronger endorsement of explicit stereotypes than women, this difference was non-significant in Japan [Germany: M_{Women} = 4.88 (4.85; 4.91), SE = 0.02, M_{Men} = 5.22 (5.18; 5.26), SE = 0.02, t (11,597) = 11.86, SE = 0.02, $p < 0.001$, d = 0.37; Japan: M_{Women} = 5.36 (5.32; 5.39), SE = 0.02, M_{Men} = 5.37 (5.33; 5.41), SE = 0.02, t (11,597) = 0.61, SE = 0.02, p = 0.541, d = 0.01]. Both men and women from Japan showed stronger stereotype endorsement than men and women from Germany [men: t (11,597) = 5.33, SE = 0.03, $p < 0.001$, d = 0.15, women: t (11,597) = 19.09, SE = 0.03, $p < 0.001$, d = 0.50]. Results were robust when year of data collection was controlled.

Discussion

In line with Study 1, Study 2 supported Hypothesis 1, showing that in large samples and with implicit and explicit stereotype measures, men were more strongly associated with science than women. It is prudent to note that significant country and gender differences should be interpreted with caution due to large sample sizes, effect sizes for country and gender differences were small to medium (0.01 < Cohen's d < 0.50). Study 2 replicated and extended findings from Study 1 and previous research (Ikkatai et al., 2019; Steffens and Jelenec, 2011), as it included much larger samples and explicit as well as implicit measures of gender-science stereotypes, whereas Study 1 focused on gender-math stereotypes. Taken together, Studies 1 and 2 take multi-faceted angles and present a comprehensive picture of gender stereotypes in the STEM domain. Based on the combined results, we

conclude that negative gender stereotypes about women's STEM ability are widespread in both countries. Study 3 thus focused on the consequences of these stereotypes and investigated to what extent female STEM students expect backlash for their stereotype-incongruent study major.

STUDY 3: EXPECTED BACKLASH FOR FEMALE STEM STUDENTS

Study 3 was a scenario study with German and Japanese university students of physics, engineering, and computer science as participants. In an online questionnaire, participants imagined being asked about their study major in a conversation with an unknown person of the opposite gender. They completed items on the expected reactions of the conversation partner, their emotions and study motivation. We hypothesized expected backlash (i.e., expected negative reactions of the conversation partner and lower ascribed communion) for women, but not for men (Hypothesis 2). This expected backlash should predict more negative/less positive emotions and lower study motivation (Hypotheses 3 and 4). Moreover, we expected these relationships to be stronger for women endorsing an interdependent self-construal (Hypothesis 5). Hypotheses were pre-registered (https://osf.io/afqxb/).

Participants and Procedure

Data were collected between January and September 2019. After registering their e-mail address in an online form, participants were invited to participate in two parts of an online questionnaire *via* personalized emails. Data from Part 1 and 2 (2-days interval between measurements) were matched with participant-generated codes. E-mail addresses could not be connected to questionnaire data. Participants provided written consent in accordance with EU General Data Protection Law. The study was approved by the ethics committee of the first author's institution.

Participants were recruited via university classes and Facebook groups/mailing lists of student associations of physics, mathematics, computer science, and engineering. The questionnaire (both parts) was completed by 656 participants. We excluded participants who were not university students or indicated non-STEM majors (n = 24), entered non-corresponding gender information at the two parts (n = 2), or indicated "other" as their gender (n = 2). The final sample consisted of 628 participants (Japanese: n = 432, 101 female, age: 18–33 years, M = 19.73, SD = 1.59; German: n = 196, 87 female, age: 18–57 years, M = 26.88, SD = 8.37).

A sample size of 100 female students per country was determined based on an a-priori power analysis for a repeated-measures ANOVA (Hypothesis 2) with a within-between interaction (medium effect size of f = 0.15, α = 0.05, power = 0.80, 2 groups, 2 measurements), which resulted in a sample size of N = 90. As Hypotheses 3–5 required path modeling, sample size was increased to 100 female students per country (and at least as many male students), resulting in a total minimum sample size of N = 400. The pre-registered sample size of female students was reached for the Japanese but not the German sample (n = 87). Data collection was terminated after 9 months of contacting Facebook groups and student councils of the STEM majors of all German universities, and 152 German university instructors. A sensitivity analysis showed that with the current sample small effects (f = 0.06) could be detected.

Materials

Materials were translated by the project team and back-translated by a professional translator. Moderators and demographics were assessed in Part 1, scenarios and outcomes in Part 2.

Part 1

Participants indicated whether they were university students, their field of study and gender. Independent/interdependent self-construal was measured with 10 items each (e.g., "I always try to have my own opinion," "I will sacrifice my self-interest for the benefit of the group I am in," 1 = do not agree, 7 = completely agree; Park and Kitayama, 2014).

Part 2

Participants were asked to imagine a conversation with an unknown person of the opposite gender in two scenarios. Female participants imagined a male conversation partner, whereas male participants imagined a female conversation partner. The wedding party scenario should represent high relational mobility (a flexible social network and opportunities to form relationships by choice), whereas the choir scenario should represent low relational mobility (a fixed network and long-term relationships due to circumstance; Thomson et al., 2018).

Wedding Party Scenario

"Please imagine you are attending a friend's wedding reception. You are introduced to a male/female person whom you have not met before. You start chatting with him/her and you feel like you are getting along well. During your conversation, he/she asks you about your university major. You tell him/her that you study (subject entered by participant displayed). Please take some time to imagine yourself in this situation."

Choir Scenario

"Please imagine that you recently decided to participate in your university's choir. Therefore, you attend the first choir meeting of the new semester. You are very motivated to join the choir and go to rehearsals regularly because you like singing and want to start a new extra-curricular activity for the next year. During the first meeting, a choir member asks you about your university major. You tell him/her that you study (subject). Please take some time to imagine yourself in this situation."

Following each scenario, participants described how they imagined the conversation partner's reaction ["How do you think would your conversation partner react to hearing that you study (subject)? Please write down his/her imagined reaction as detailed as possible. Keep in mind that reactions can either be verbal (what he/she says) or non-verbal (facial

TABLE 1 | Descriptive statistics.

		Reaction	Communion	Positive emotions	Negative emotions	Anxiety emotions	Motivation	Independent self-construal
Japan								
Men	*M*	4.42	3.97	3.71	2.27	3.32	4.43	4.41
	[95% CI]	[4.31; 4.52]	[3.84; 4.08]	[3.58; 3.83]	[2.15; 2.39]	[3.19; 3.45]	[4.32; 4.53]	[4.32; 4.50]
(*n* = 331)	α	0.74	0.89	0.92	0.82	0.76	0.86	0.68
Women	*M*	4.48	3.41	3.64	2.09	2.97	4.29	4.43
	[95% CI]	[4.30; 4.68]	[3.20; 3.60]	[3.44; 3.83]	[1.87; 2.33]	[2.71; 3.23]	[4.11; 4.47]	[4.28; 4.58]
(*n* = 101)	α	0.85	0.88	0.88	0.87	0.81	0.85	0.68
Germany								
Men	*M*	4.69	4.03	4.24	1.85	2.18	4.22	4.74
	[95% CI]	[4.53; 4.85]	[3.78; 4.28]	[4.01; 4.46]	[1.67; 2.04]	[1.96; 2.41]	[4.12; 4.34]	[4.60; 4.86]
(*n* = 109)	α	0.72	0.92	0.92	0.89	0.80	0.76	0.63
Women	*M*	4.87	4.24	4.67	2.13	2.14	4.49	4.89
	[95% CI]	[4.68; 5.07]	[3.97; 4.50]	[4.47; 4.89]	[1.91; 2.38]	[1.91; 2.38]	[4.29; 4.70]	[4.74; 5.04]
(*n* = 87)	α	0.59	0.94	0.89	0.92	0.83	0.92	0.67
Total								
Men	*M*	4.49	3.98	3.84	2.17	3.04	4.38	4.49
	[95% CI]	[4.40; 4.58]	[3.88; 4.09]	[3.72; 3.95]	[2.06; 2.27]	[2.92; 3.16]	[4.29; 4.46]	[4.42; 4.56]
(*n* = 440)	α	0.74	0.90	0.92	0.88	0.80	0.84	0.67
Women	*M*	4.66	3.79	4.12	2.11	2.58	4.38	4.64
	[95% CI]	[4.53; 4.81]	[3.62; 3.97]	[3.95; 4.29]	[1.95; 2.28]	[2.40; 2.77]	[4.26; 4.51]	[4.53; 4.75]
(*n* = 188)	α	0.71	0.92	0.91	0.90	0.83	0.88	0.68

Note: For scales with more than two items, Chronbach's α is displayed, for reaction and motivation Spearman's rho is displayed.

expression, body language etc.)."], rated the reaction valence ("How positive or negative is this reaction?" 1 = very negative, 7 = very positive) and impression ("How positive or negative do you think is your conversation partner's impression of you?" 1 = very negative, 7 = very positive).

Furthermore, they rated expected communion ("Please indicate how much your conversation partner thinks you possess the following traits," 4 items, gentle, affectionate, supportive, sympathetic; Steinmetz et al., 2014; 1 = not at all, 7 = completely), emotions ("How would you feel in the scenario?" 13 items; Morinaga et al., 2017, 1 = do not agree, 7 = completely agree), and motivation ["In the scenario, how would you intend to work hard for (subject) from now on? Please indicate whether your motivation is stronger or weaker compared to before." 1 = much weaker than before, 7 = much stronger than before; and "In the scenario, how has your motivation to study hard for (subject indicated above) changed?" 1 = completely lost motivation; 7 = motivation got much stronger; Morinaga et al., 2017]. Demographics included field of study, gender, birth year, and nationality. Further measures not reported in this paper were implicit theories of intelligence, gender identity, implicit gender-science attitudes, and benevolent sexism (Part 1), perceived agency, general motivation, career and research intentions, goals, perceived stereotype threat, future work domain and importance of digitalization for STEM (Part 2).

Statistical Analyses

Because all questions were programmed as mandatory in the online questionnaire, there was no missing data. Measurement invariance was tested with exploratory factor analysis (conducted in SPSS version 25) and confirmatory factor analysis (conducted in Mplus Version 8.6). Cutoff criteria for goodness of model fit in CFA were CFI/TLI ≥0.90, SRMR ≤0.06, RMSEA ≤0.08. Reaction valence and communion stereotypes were investigated with linear mixed models with REML estimation. Open-ended answers on reactions were categorized and subjected to frequency analysis (cross tabulation and χ^2 tests). Consequences of reactions and communion for emotions and motivation were investigated with path analysis in Mplus.

Results

Measurement Invariance and Descriptive Statistics

We investigated measurement invariance between national subsamples for multi-item measures. Multiple-group confirmatory factor analysis (CFA) with the national groups after model modifications showed partial metric invariance for all scales. For emotions, exploratory factor analysis (EFA) with promax rotation yielded three factors: positive (happy, proud, feeling good, satisfied, relieved, relaxed), negative (disappointed, angry, feeling bad, dissatisfied), and anxiety (anxious, nervous, embarrassed). In a CFA configural model (no equality constraints), three items (satisfied, relieved, feeling bad) were excluded due to low factor loadings and high cross-loadings. Loadings of item "angry" on negative emotions and item "embarrassed" on anxiety emotions were freed due to non-equivalence. The model showing partial metric invariance (i.e., factor structure and at least two loadings per factor constrained to be equal across groups) had acceptable model

TABLE 2 | Bivariate correlations [r, (95% CI)].

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Germany							
(1) Reaction	—	0.40*** [0.23; 0.55]	0.69*** [0.57; 0.78]	−0.46*** [−0.61; −0.31]	−0.35*** [−0.52; −0.14]	0.37*** [0.19; 0.52]	0.09 [−0.09; 0.26]
(2) Communion	0.36** [0.12; 0.53]	—	0.51*** [0.32; 0.67]	−0.13 [−0.29; 0.02]	−0.00 [−0.20; 0.18]	0.19 [0.02; 0.35]	0.11 [−0.08; 0.29]
(3) Positive emotions	0.67*** [0.49; 0.78]	0.35** [0.14; 0.52]	—	−0.25** [−0.46; −0.05]	−0.23* [−0.42; 0.00]	0.36*** [0.21; 0.48]	0.14 [−0.06; 0.33]
(4) Negative emotions	−0.48*** [−0.67; −0.22]	−0.23* [−0.44; 0.00]	−0.48*** [−0.64; −0.27]	—	0.60*** [0.44; 0.73]	−0.22* [−0.42; −0.02]	0.04 [−0.18; 0.23]
(5) Anxiety emotions	−0.24* [−0.42; −0.05]	0.04 [−0.16; 0.22]	−0.45*** [−0.57; −0.30]	0.58*** [0.41; 0.73]	—	−0.15 [−0.35; 0.06]	−0.10 [−0.27; 0.08]
(6) Motivation	0.14 [−0.19; 0.45]	0.01 [−0.26; 0.31]	0.26* [−0.06; 0.52]	0.01* [−0.24; 0.29]	−0.15 [−0.30; 0.03]	—	0.00 [−0.20; 0.21]
(7) Independent self-construal	0.09 [−0.13; 0.31]	0.08 [−0.18; 0.32]	0.24* [−0.01; 0.46]	−0.09 [−33; 0.15]	−0.37*** [−0.55; −0.17]	0.31** [0.08; 0.50]	—
Japan							
(1) Reaction	—	0.45*** [0.33; 0.55]	0.61*** [0.51; 0.69]	−0.33*** [−0.44; −0.22]	−0.11 [−0.23; 0.02]	0.60*** [0.50; 0.69]	−0.01 [−0.12; 0.08]
(2) Communion	0.23* [0.02; 0.42]	—	0.63*** [0.53; 0.71]	0.05 [−0.06; 0.17]	0.20*** [0.07; 0.33]	0.42*** [0.30; 0.53]	−0.07 [−0.17; 0.04]
(3) Positive emotions	0.51*** [0.33; 0.66]	0.61*** [0.47; 0.74]	—	−0.01 [−0.11; 0.12]	0.14* [0.01; 0.28]	0.48*** [0.35; 0.59]	−0.07 [−0.18; 0.04]
(4) Negative emotions	−0.61*** [−0.73; −0.44]	−0.01 [−0.21; 0.26]	−0.27** [−0.47; 0.03]	—	0.51*** [0.43; 0.58]	−0.25*** [−0.36; −0.13]	−0.11 [−0.22; 0.03]
(5) Anxiety emotions	−0.25* [−0.41; −0.05]	0.20* [−0.03; 0.41]	0.15 [−0.08; 0.36]	0.51*** [0.38; 0.64]	—	−0.01 [−0.14; 0.11]	−0.24*** [−0.35; −0.13]
(6) Motivation	0.64*** [0.45; 0.76]	0.36*** [0.17; 0.52]	0.58*** [0.44; 0.70]	−0.48*** [−0.66; −0.24]	−0.05 [−0.24; 0.16]	—	−0.05 [−0.06; 0.17]
(7) Independent self-construal	−0.06 [−0.24; 0.13]	0.01 [−0.23; 0.21]	0.02 [−0.21; 0.23]	−0.09 [−0.34; 0.13]	−0.20 [−0.39; 0.01]	0.09 [−0.12; 0.30]	—

Notes: Correlations for women (men) below (above) the diagonal.
*p < *0.05*, **p < *0.01*, ***p < *0.001*.

fit [χ^2 (72) = 218.58, $p < 0.001$, RMSEA = 0.08, CFI = 0.96, TLI = 0.95, SRMR = 0.06] which did not significantly differ from the configural model [$\Delta\ \chi^2$ (8) = 9.31, $p = 0.317$].

EFA and CFA showed that the four communion items loaded on a single factor in both samples. A model showing partial metric invariance showed good fit when error terms of the items "supportive" and "sympathetic" were allowed to correlate, and the loading of the item "compassionate" was freed due to non-invariance, [χ^2 (3) = 3.82, $p = 0.282$, RMSEA = 0.03, CFI = 1.00, TLI = 1.00, SRMR = 0.07]. The fit of this model did not significantly differ from the configural model [$\Delta\ \chi^2$ (1) = 0.73, $p = 0.393$].

Internal consistency of the interdependent self-construal scale was not acceptable (0.55 < Cronbach's α < 0.72) with German men showing a value below 0.60, therefore, we used the independent self-construal scale, which had acceptable consistency (0.63 < α < 0.68). Issues with low reliability of this and other self-construal scales have been reported and discussed in earlier research (e.g., Gudykunst and Lee, 2003; Park and Kitayama, 2014). Multi-group CFA showed partial metric invariance in a single-factor model when error terms of "I always try to have my own opinions" and "I always express my opinions clearly" were allowed to correlate, and the loadings of the items "I always try to have my own opinions" and "It does not concern me when my opinions or behavior differs from that of other people" were freed due to non-invariance [χ^2 (74) = 136.08, $p < 0.001$, RMSEA = 0.05, CFI = 0.92, TLI = 0.90, SRMR = 0.05]. The fit of this model did not significantly differ from the configural model [$\Delta\ \chi^2$ (6) = 11.87, $p = 0.065$]. Independent and interdependent self-construal were negatively correlated in all subgroups ($-0.12 < r < -0.54$). Descriptive statistics are displayed in **Table 1**, bivariate correlations in **Table 2**. Outputs of CFAs to investigate measurement invariance can be found on the OSF.

Reaction Valence and Communion Stereotypes

To test Hypothesis 2, data were transformed into long format (1,256 observations, 628 participants) due to the repeated measurements for the scenarios (in long format, one data row represented one observation instead of one participant). We computed linear mixed models with Gender (male vs. female, between-participants) and Scenario (high vs. low relational mobility, within-participants) as factors and valence of imagined reactions and communion as dependent variables. For reaction valence, there was a main effect of Gender, F (1, 1,251.97) = 6.22, $p = 0.013$. Women expected more positive reactions [$M = 4.66$ (4.55; 4.78), $SE = 0.04$] than men [$M = 4.49$ (4.41; 4.56), $SE = 0.06$, t (1,251.97) = 2.50, $SE = 0.07$, $p = 0.013$, $d = 0.15$].

TABLE 3 | Frequencies of categories of reactions to scenarios.

	Positive	Negative	Surprised	Neutral	Total
High relational mobility					
male	191 (43%)	69 (16%)	94 (21%)	85 (20%)	439 (100%)
female	39 (21%)	10 (5%)	123 (65%)	16 (9%)	188 (100%)
Low relational mobility					
male	159 (36%)	72 (17%)	80 (18%)	128 (29%)	439 (100%)
female	45 (24%)	16 (9%)	80 (43%)	47 (25%)	188 (100%)

The main effect of Scenario was also significant, F (1, 1,251.97) = 15.32, $p < 0.001$. Valence of reactions was more positive in the high relational mobility scenario [M = 4.71 (4.62; 4.81), SE = 0.05] compared to the low relational mobility scenario [M = 4.44 (4.34; 4.53), SE = 0.05, t (1,251.97) = 3.92, SE = 0.07, $p < 0.001$, d = 0.22]. The interaction was non-significant [t (1,251.97) = 0.81, p = 0.368].

For communion, there was a main effect of Gender, F (1, 1,247.91) = 6.36, p = 0.012. Women expected lower communion ratings [M = 3.79 (3.67; 3.92), SE = 0.06] than men [M = 3.98 (3.90; 4.07), SE = 0.04, t (1,247.91) = 2.53, SE = 0.08, p = 0.012, d = 0.22]. The main effect of Scenario and the interaction were non-significant (Fs < 1.46, ps > 0.227).

In an exploratory analysis, we categorized the open-ended responses on imagined reactions into four categories: positive, negative, surprised, and interested/neutral. Reactions were coded as positive when containing positive aspects (e.g., impressed, admiring, interested) and negative when containing negative aspects (e.g., rejection, disinterest, distancing, negative comments about STEM). Reactions were coded as surprised when cues for surprise were mentioned (e.g., surprised, amazed, perplexed). When surprise was mentioned in combination with other aspects, responses were coded as surprised. Reactions were coded as neutral when they did not contain positive, negative or surprised aspects (e.g., no apparent reaction, neutral) or if a combination of positive and negative reactions was mentioned (e.g., "it could be one of two possibilities, a positive reaction or a negative one"). Reactions were coded by two independent raters each who were fluent in the respective language. Interrater reliabilities were excellent (Cohen's Kappa ≥ 0.89), indicating high agreement between raters. There were gender differences in distributions across categories: men more frequently described positive expected reactions, whereas women more frequently described surprised reactions [**Table 3**, high relational mobility: χ^2 (3) = 113.19, $p < 0.001$; low relational mobility: χ^2 (3) = 43.29, $p < 0.001$]. Further descriptive analyses showed that none of the men, but 34% of women imagined their conversation partner to make a reference to their gender [e.g., "Really? Women are rare in (subject)," "理系女" (Rikejo, female scientist), "you are studying (subject) although you are a woman?"].

Consequences of Perceived Reactions and Communion

To test Hypotheses 3 and 4, we computed path models with multiple-group comparison for men/women including reaction valence and communion ratings as predictors of emotions. In turn, reactions and emotions predicted study motivation. For constructs measured with more than two items (i.e., communion, emotions, self-construal) we used factor scores that were generated under the assumption of partial metric invariance as manifest variables in the model (please note that we deviate from the pre-registration which included latent variable modeling to account for the fact that partial metric invariance but not full scalar invariance could be established). Scenarios were combined, but separate analyses showed similar results with larger effect sizes for the low relational mobility scenario. Results for separate analyses can be found on the OSF. Although the pre-registration stated that we would use scenario as a covariate, we opted for presenting the results for the scenarios separately to better reflect potential differences between scenarios. We controlled for country of data collection in all analyses.

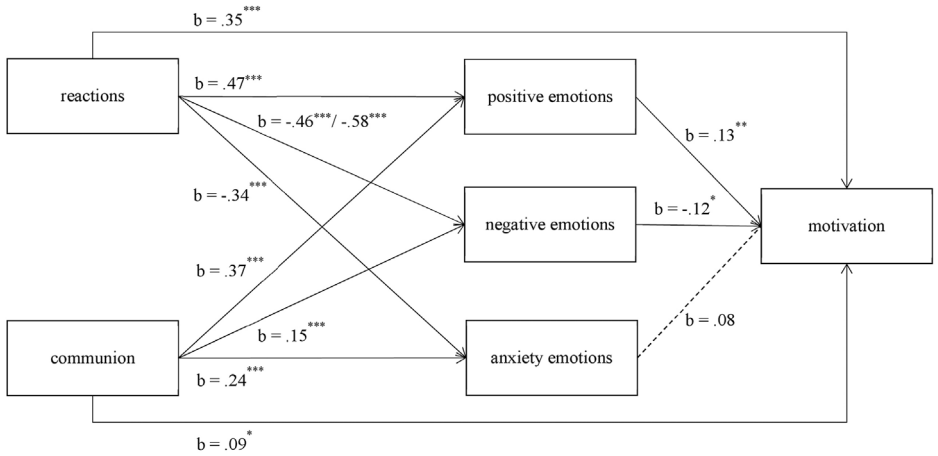
In a first model, all paths were constrained to be equal for men and women. Country of data collection was entered as a control

TABLE 4 | Direct effects in modified path model.

	b	[LLCI; ULCI]	SE	p
Reaction				
country	−0.34	[−0.51; −0.17]	0.09	<0.001
Communion				
country	−0.03	[−0.20; 0.14]	0.09	0.753
Positive emotions				
reaction	0.47	[0.41; 0.53]	0.03	<0.001
communion	0.37	[0.31; 0.43]	0.03	<0.001
country	0.15	[0.03; 0.27]	0.06	0.011
Negative emotions				
reaction (men)	−0.46	[−0.54; −0.38]	0.03	<0.001
reaction (women)	−0.58	[−0.68; −0.47]	0.05	<0.001
communion	0.15	[0.07; 0.22]	0.04	<0.001
country	−0.17	[−0.32; −0.02]	0.08	0.028
Anxiety emotions				
reaction	−0.34	[−0.42; −0.27]	0.04	<0.001
communion	0.24	[0.16; 0.31]	0.04	<0.001
country	−0.14	[−0.29; 0.01]	0.08	0.069
Motivation				
reaction	0.35	[0.26; 0.44]	0.04	<0.001
communion	0.09	[0.01; 0.16]	0.04	0.031
positive emotions	0.13	[0.04; 0.22]	0.05	0.003
negative emotions	−0.12	[−0.22; −0.02]	0.05	0.022
anxiety emotions	0.08	[−0.02; 0.18]	0.05	0.125
country	0.19	[0.06; 0.33]	0.07	0.006

TABLE 5 | Indirect effects in modified path model.

	ab	[LLCI; ULCI]	SE	p
Reaction → motivation				
positive emotions	0.06	[0.02; 0.10]	0.02	0.004
negative emotions (men)	0.06	[0.01; 0.10]	0.02	0.025
negative emotions (women)	0.07	[0.01; 0.13]	0.03	0.025
anxiety emotions	−0.03	[−0.06; 0.01]	0.02	0.130
Communion → motivation				
positive emotions	0.05	[0.02; 0.08]	0.02	0.004
negative emotions	−0.02	[−0.04; 0.00]	0.01	0.049
anxiety emotions	0.02	[−0.01; 0.04]	0.01	0.137



Notes. Dashed arrow indicates non-significant path. Path from reactions to negative emotions: coefficients for men/women.

FIGURE 1 | Results of the modified path model (H3 and H4).

variable. Model fit was good [χ^2 (17) = 38.20, p = 0.002, RMSEA = 0.064, CFI = 0.99, TLI = 0.97, SRMR = 0.05]. Modification indices showed that fit could be further improved by relaxing the constraint for the path from reactions to negative emotions. Constraints were relaxed for the direct and indirect effects. Fit of the modified model was good [χ^2 (16) = 33.51, p = 0.006, RMSEA = 0.06, CFI = 0.99, TLI = 0.97, SRMR = 0.05]. A χ^2 difference test showed that the fit of this modified model was significantly better than that of the fully constrained model [$\Delta\chi^2$ (1) = 4.69, p = 0.030]. Results are displayed in **Table 4** (direct effects), **Table 5** (indirect effects), and **Figure 1**. All direct and indirect effects were in the expected direction and significant, except for a non-significant path from anxiety emotions to motivation and the indirect effects *via* anxiety emotions, which were both non-significant. The path from reactions to negative emotions and the indirect effect of reactions to motivation *via* negative emotions were stronger for female than male students. Unexpectedly, the paths from reactions and communion to positive and anxiety emotions were equal for female and male students.

Moderation by Self-Construal

To investigate whether self-construal moderated the relationships from reactions and communion to emotions and motivation (Hypothesis 5), we introduced independent self-construal and its interactions with reactions, communion, and emotions as additional predictors, controlling for country. Predictors involved in the interactions were centered. The additional paths were unconstrained. Model fit was not acceptable [χ^2 (64) = 141.94, p < 0.001, RMSEA = 0.06, CFI = 0.95, TLI = 0.92, SRMR = 0.07]. Inspection of results showed that independent self-construal interacted with reactions and communion to predict emotions, but did not interact with emotions to predict motivation. Thus, we computed a modified model excluding the interactions of emotions and self-construal on motivation. This modified model had acceptable fit [χ^2 (34) = 71.58, p < 0.001, RMSEA = 0.06, CFI = 0.98, TLI = 0.95, SRMR = 0.06]. For male participants, there was an interaction of communion and independent self-construal to predict negative emotions as well as anxiety emotions. Simple slopes analyses showed that for male participants weakly endorsing an independent self-construal, higher communion ratings predicted higher negative emotions (b = 0.28, *SE* = 0.06, p < 0.001) and higher anxiety emotions (b = 0.38, *SE* = 0.06, p < 0.001), whereas these relationships were non-significant for male participants strongly endorsing and independent self-construal (negative emotions: b = 0.03, *SE* = 0.06, p = 0.594; anxiety emotions: b = 0.10, *SE* = 0.06, p = 0.121) Moreover, self-construal interacted with reactions to predict positive emotions. For male participants weakly endorsing an independent self-construal, the relationship of positive reactions and positive emotions was stronger (b = 0.38, *SE* = 0.05, p < 0.001) than for male participants strongly endorsing an independent self-construal (b = 0.28, *SE* = 0.05, p < 0.001). In turn, for female participants there were no direct or moderated effects of self-construal on positive or negative emotions, but self-construal negatively predicted anxiety emotions and positively predicted motivation. Results are depicted in **Table 6**.

Discussion

Study 3 showed mixed evidence for expected backlash for women in STEM (Hypothesis 2). Women expected their conversation partner to react more positively than men (contrary to expectations), but expected to be rated lower in communion than men (in line with expectations). These results might indicate a subtle expected backlash effect in that female students did not imagine blatant negative reactions to disclosing their STEM major to the conversation partner, but they expected lower communion ratings. This latter result is consistent with lack-of-fit models indicating that women in agentic fields (in the United States) are rated lower in communion for disconfirming the female stereotype (Rudman and Phelan, 2008).

TABLE 6 | Results of path model with moderation by independent self-construal.

	b	[LLCI; ULCI]	SE	p
Reaction				
country	−0.34	[−0.51; −0.17]	0.09	<0.001
Communion				
country	−0.03	[−0.20; 0.14]	0.09	0.754
Independent self-construal				
country	0.01	[−0.14; 0.16]	0.08	0.906
Positive emotions				
reaction	0.47	[0.41; 0.53]	0.03	<0.001
communion	0.37	[0.31; 0.43]	0.03	<0.001
self-construal (men)	−0.03	[−0.10; 0.05]	0.04	0.515
reaction*self-construal (men)	−0.10	[−0.19; −0.02]	0.04	0.020
communion*self-construal (men)	0.07	[−0.02; 0.16]	0.05	0.122
self-construal (women)	0.05	[−0.06; 0.16]	0.06	0.403
reaction*self-construal (women)	0.02	[−0.10; 0.13]	0.06	0.802
communion*self-construal (women)	0.02	[−0.10; 0.14]	0.06	0.769
country	0.14	[0.04; 0.28]	0.06	0.015
Negative emotions				
reaction (men)	−0.46	[−0.54; −0.38]	0.04	<0.001
reaction (women)	−0.55	[−0.66; −0.44]	0.06	<0.001
communion	0.15	[0.08; 0.23]	0.04	<0.001
self-construal (men)	−0.11	[−0.20; −0.01]	0.05	0.023
reaction*self-construal (men)	0.06	[−0.05; 0.16]	0.05	0.300
communion*self-construal (men)	−0.14	[−0.25; −0.03]	0.06	0.012
self-construal (women)	−0.12	[−0.26; 0.02]	0.07	0.102
reaction*self-construal (women)	−0.12	[−0.27; 0.04]	0.08	0.149
communion*self-construal (women)	−0.02	[−0.18; 0.13]	0.08	0.755
country	−0.17	[−0.32; −0.02]	0.08	0.023
Anxiety emotions				
reaction	−0.34	[−0.42; −0.27]	0.04	<0.001
communion	0.24	[0.16; 0.31]	0.04	<0.001
self-construal (men)	−0.22	[−0.31; −0.13]	0.05	<0.001
reaction*self-construal (men)	0.04	[−0.06; 0.15]	0.05	0.416
communion*self-construal (men)	−0.16	[−0.27; −0.05]	0.06	0.004
self-construal (women)	−0.26	[−0.40; −0.11]	0.07	0.001
reaction*self-construal (women)	−0.05	[−0.21; 0.10]	0.08	0.509
communion*self-construal (women)	−0.03	[−0.18; 0.12]	0.08	0.701
country	−0.14	[−0.30; 01]	0.08	0.058
Motivation				
reaction	0.35	[0.27; 0.44]	0.04	<0.001
communion	0.08	[0.00; 0.16]	0.04	0.038
self-construal (men)	0.06	[−0.02; 0.14]	0.04	0.171
self-construal (women)	0.19	[0.05; 0.32]	0.07	0.008
positive emotions	0.13	[0.04; 0.22]	0.05	0.004
negative emotions	−0.13	[−0.23; −0.03]	0.05	0.012
anxiety emotions	0.11	[0.01; 0.21]	0.05	0.040
country	0.19	[0.06; 0.33]	0.07	0.005

Contrary to Hypotheses 3 and 4, the path models largely showed gender similarities. However, results might also point to some negative consequences of the (subtle) expected backlash effect: Less positive reactions were related to more negative emotions (and consequently lower motivation) more strongly for female than male students. This might indicate that (some) women are sensitive to disconfirming the female stereotype and consequently suffer negative consequences of backlash. Concerning Hypothesis 5, results indicated that self-construal played a moderating role for male, but not for female participants. Lower independent self-construal was associated with stronger relationships of reactions and communion to emotions than higher independent self-construal. This result might indicate that men who see themselves as less independent from social others are more susceptible to possible positive and negative effects of social reactions on their emotions. Whereas positive reactions were related to positive emotions, higher communion was related positive, as well as to negative and anxiety emotions. This pattern of results might represent a double-edged sword of communion for men who place less value on being independent of social others: on the one hand, it is in general socially desirable to be rated high on communion (Steinmetz et al., 2014), on the other hand being perceived as highly communal might induce masculinity threat due to precarious manhood beliefs (e.g., Bosson et al., 2021; Vandello and Bosson, 2013). Unexpectedly, self-construal did not play a moderating role for female participants, indicating that for women, self-construal did not relate to a higher or lower susceptibility to expected backlash effects.

GENERAL DISCUSSION

Women are underrepresented in STEM fields like physics, engineering and computer science around the world, including Germany and Japan, which are top-ranking in STEM (e.g., Nature Index, 2020; U. S. News and World Report LP, 2020). Whereas there are no gender differences in math ability (e.g., Else-Quest et al., 2010; Lindberg et al., 2010), stereotypes play a role in gender segregation in STEM (e.g., Nosek et al., 2009). Based on social role theory (Eagly and Wood, 2012) and role congruity theory (Eagly and Karau, 2002), we expected gender-science stereotypes to be associated with expected backlash (Rudman and Phelan, 2008) for female STEM students, which negatively affects their emotions and motivation.

Studies 1 and 2 showed that widespread gender stereotypes in Germany and Japan associated men with math and science and women with liberal arts (Steffens and Jelenec, 2011). Results were consistent when using mixed methods including open-ended questions, Likert-scale explicit stereotype measurement as well as the Implicit Association Test, spanning multiple years of measurement (2006–2017). Replicating previous research that was conducted in each country separately (Ikkatai et al., 2020; Steffens and Jelenec, 2011), the current research showed in a joint investigation of both countries that in line with Hypothesis 1, negative stereotypes about women's STEM ability were endorsed in both Germany and Japan. We thus conclude that these stereotypes likely contribute to women's underrepresentation in STEM in these countries.

Study 3 investigated expected backlash as a potential consequence of gender-science stereotypes. A scenario study with students of physics, engineering, and computer science from Germany and Japan as participants showed tentative evidence for expected backlash for female STEM students.

Concerning Hypotheses 2–4, gender differences were not as clear and pronounced as expected. Associations between expected reactions and communion to emotions and motivation were largely similar for male and female STEM students. Nevertheless, results point to subtle expected backlash effects for female students: They expected their conversation partner to rate them lower on communion (but not to react more negatively) than male participants. Furthermore, they more frequently expected surprised reactions than men, and 34% of women (0% of men) imagined their conversation partner to refer to their gender in the reactions. We take this as evidence that for women, gender is more salient in the imagined conversation about their study major. This salience might indicate that studying a STEM subject is seen as counter-stereotypical behavior violating prescriptive gender stereotypes. Female STEM students might thus expect that others perceive a lack of fit of women to STEM (Heilman, 1983; Eagly and Karau, 2002).

Results of Study 3 imply expected backlash for women in STEM, but participants did not expect this backlash to be blatantly negative. Backlash can manifest itself in subtle emotional responses like frowning or derisive smiling, which are discussed as possible indicators of implicit social punishment for disconfirming gender stereotypes (Rudman and Phelan, 2008). Such subtle responses can also be conceptualized as micro-aggressions (i.e., "brief, everyday exchanges that send denigrating messages to individuals because of their group membership," Sue, 2010, p. xvii). Gender-based micro-aggressions in STEM contexts have recently gained attention (e.g., Sekaquaptewa, 2019) and might have contributed to subtle expected backlash. Moreover, the associations from less positive reactions to motivation via negative emotions were significantly stronger for female than for male students. This result indicates that even a subtle expected backlash might have negative consequences for female STEM students.

The cross-cultural approach showed that gender-science stereotypes were endorsed in Germany and Japan to a similar extent, corresponding to the comparable underrepresentation of women in STEM in these countries. Furthermore, associations between expected backlash, emotions and motivation remained consistent when country of data collection was statistically controlled for. A model investigating independent self-construal as a moderator showed two noteworthy patterns of results. First, self-construal moderated the paths from reactions/communion to emotions, but not the paths from emotions to motivation. This might indicate that self-construal is more relevant for how social reactions are perceived and which emotions are elicited by these perceptions. In turn, these emotions were associated with motivation to study irrespective of the level of self-construal endorsed, speaking for effects of emotions on motivation that are generalizable across participants' cultural orientations. Second, self-construal moderated paths from reactions/communion to emotions only for male, but not for female participants. Thus, Hypothesis 5, that individuals endorsing an independent self-construal are less prone to expecting negative social repercussions for counter-stereotypical behavior of studying a STEM subject, was only supported for male participants. In contrast, female participants were susceptible to consequences of expected reactions to studying a STEM subject irrespective of the relevance of the group for their self. However, it should be noted that reliabilities of self-construal were at the lower end and unsatisfactory for some groups, calling for a replication with more reliable measures of self-construal.

Results were similar across scenarios, but stronger for the scenario representing low relational mobility. This indicates that the experienced negative consequences of stereotypes might be stronger for women in STEM in contexts in which the social network is more stable and less based on personal choice, as social rejection has more severe consequences in these contexts (Sato et al., 2014). Future studies should substantiate this preliminary evidence that the intensity of consequences of expected backlash for stereotype-incongruent behavior of women in STEM might be aggravated by the cultural factor of relational mobility.

Theoretical and Practical Implications

The present research showed that in the STEM fields of physics, engineering and computer science, similar social-psychological mechanisms as in other male-dominated domains (e.g., leadership) might impede gender equality in Germany and Japan. In accordance with social role theory (Eagly and Wood, 2012), the observation of gender segregation in male-dominated STEM fields is associated with the stereotype that in Germany and Japan, men are stereotypically perceived as better-suited for STEM than women. We applied role congruity theory (Eagly and Karau, 2002) to the STEM context to shed light on the psychological processes contributing to women's underrepresentation in STEM in Germany and Japan. Like in the leadership domain, women might experience backlash effects in gender-segregated STEM fields. Because STEM is incongruent with the traditional female social role, women in STEM might experience social rejection in cultural contexts like Germany and Japan, where negative gender-science stereotypes are widespread (Study 2, but see also Ikkatai et al., 2020; Steffens and Jelenec, 2011).

A recent model advancing role congruity theory describes the interplay of social roles and motivational causes for gender inequality in STEM. Goal congruity theory (Diekman et al., 2017) posits that gender roles build an opportunity structure to fulfill individual (stereotype-congruent) goals. Women tend to strive for communal goals (e.g., helping other people), whereas men tend to strive for agentic goals (e.g., gaining power). By valuing different goals, women and men select into stereotype-congruent roles (study fields and careers). STEM fields are not perceived as affording communal goals. Thus, pursuing STEM creates goal incongruity for women, which can lead to lower motivation and opting out of STEM. The current research showed that women's motivation is impaired by expected negative social reactions to studying a STEM subject. These negative reactions as a signal of lack-of-fit might communicate to women that STEM is perceived as incongruent with their gender role, thereby creating or aggravating goal incongruity. Importantly, participants in Study 3 had already successfully entered STEM majors, which means that they had sufficiently positive initial beliefs about STEM to enroll in this kind of major.

The current research shows that even for these participants, who are highly invested in pursuing a STEM major and potential career, expected negative social reactions can have detrimental consequences for their emotions and motivation. Paired with the widespread gender stereotypes, these consequences can further aggravate gender segregation in STEM, as the leaky pipeline shows that female STEM students are often less inclined to pursue a STEM career than their male counterparts (e.g., Diekman et al., 2017; Jasko et al., 2020).

Goal congruity theory might also explain why we found positive relationships of communion to positive, negative, and anxiety emotions for male as well as female students. Communal goals represent the basic need of relatedness and are thus important to everyone (Diekman et al., 2017). Moreover, communion is socially valued (e.g., Abele et al., 2008). Therefore, being perceived as communal in the scenarios was associated with more positive emotions for both genders. However, because STEM is perceived as incongruent with communal goals, goal conflict is likely to arise. Goal conflict can elicit anxiety and negative emotions (Gray and McNaughton, 2003), potentially explaining why communion was related to higher anxiety and negative emotions for both genders.

Results open up pathways to reduce women's underrepresentation in these STEM fields. Studies 1 and 2 showed that gender-science stereotypes are pervasive in Germany and Japan. There have been efforts to develop educational programs to reduce gender stereotypes and their effects, for example, focusing on teaching students a growth mindset or motivational and strategic trainings (e.g., Law et al., 2021; Moè, 2021). As changing stereotypes has been shown to be quite difficult (Heilman and Caleo, 2018), another fruitful road to gender equality in STEM in Germany and Japan is to reduce role and goal incongruity. An intervention to change communal goal affordances (i.e., the opportunities for goal pursuit) in STEM (Belanger et al., 2020) showed that perceiving communal goal affordances (e.g., collaborative lab activities) in STEM increased social belonging and interest, especially for women. Highlighting STEM's potential to afford communal goals might therefore alleviate goal incongruity and reduce gender-science stereotypes and backlash effects for women, because STEM is perceived as less incongruent with the female gender role.

Limitations and Future Directions

A first limitation of the current research is the measurement of social reactions in Study 3. Although we used a combination of open-ended and Likert-scale questions and participants were asked to imagine both verbal and non-verbal reactions, the items captured rather blatant than subtle reactions. Future research should investigate a broader variety of reactions to disconfirming stereotypes in STEM. Second, Study 3 measured backlash and its consequences only cross-sectionally. As goal incongruity might be anticipated and repeatedly experienced before it has detrimental consequences for women's STEM motivation (Diekman et al., 2017), future research should investigate consequences of backlash and incongruity in longitudinal studies. In addition, we compared two countries with similar gender segregation in STEM (e.g., Destatis, 2019; Gender Equality Bureau Cabinet Office, 2017), but different cultural orientations (self-construal, relational mobility; e.g., Markus and Kitayama, 1991; Thomson et al., 2018). To fully unfold the possible interplay of social roles, gender, and culture for women in STEM, future research should investigate a larger sample of countries with varying positions on the individualistic/post-materialistic vs. collectivistic/materialistic continuum (Yalcinkaya and Adams, 2020) as well as varying levels of gender inequality, as previous cross-cultural research conducted in different European countries (e.g., Italy, Norway, Poland, Spain, United Kingdom) has shown that the extent and consequences of gender stereotypes may be in part shaped by a country's gender inequality (e.g., Castaño et al., 2020; Bedyńska et al., 2021; Moè et al., 2021).

Third, internal consistency of self-construal was low and relational mobility was not measured on the individual level, but varied in two scenarios. The inclusion of further scales to measure these cultural variables and further moderators and mediators (e.g., perceived goal conflict) could illuminate individual factors increasing women's susceptibility to backlash and role/goal incongruity. Another possible mediator could be rejection sensitivity (Sato et al., 2014), which might explain why gender was more salient in the scenarios for some of the female participants.

Fourth, similar to many other cross-cultural studies, we were not able to establish full scalar invariance of the multi-item measures used in Study 3. The level of partial metric invariance was reached in that factor loadings of at least two items per construct were equal. This enabled us to test relationships between variables in path analysis. However, we note that intercepts were not equal between national subsamples and we therefore refrained from estimating latent variables in structural equation modeling.

Finally, the scenarios might have activated occupational stereotypes along with gender stereotypes. People–particularly men–in STEM are stereotyped to be "socially awkward" (Cheryan et al., 2013). Therefore, in addition to gender stereotypes, occupational stereotypes about social skills might have been activated. These stereotypes might have caused men to also expect backlash to their study major. Future research should therefore disentangle backlash due to gender and occupational stereotypes. As in the current study scenarios were limited to social interactions outside of STEM, it might be also worthwhile to investigate expected backlash effects in further scenarios that are related to the academic/work domain.

CONCLUSION

Factors explaining gender inequality in STEM are manifold. The present research adds to the literature by investigating social-psychological and cultural mechanisms to relatively low STEM motivation for women in Germany and Japan. A mixed-methods

investigation of gender-science stereotypes confirmed negative stereotypes about women in STEM in both countries. Even though Germany and Japan differ in cultural orientations, the impact of stereotypes on gender segregation in STEM seems to be pervasive in both countries. Recent promising measures to reduce gender inequality do not focus on changing women's individual predictors of STEM success, but rather investigate how STEM is stereotyped. As stereotypes are socially and culturally shared, cross-cultural research may further illuminate the social context of gender inequality in STEM.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Open Science Framework, https://osf.io/4awqe/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the research ethics commissioner of the FernUniversität in Hagen. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LF, ST, YM, KS, YU, GT, SM, MK, and SS designed the studies, LF, ST, YM, KS, and YU collected the data, LF analyzed the data, LF wrote the manuscript with contributions from ST, YM, KS, YU, GT, SM, MK, and SS, LF revised the manuscript with contributions from SM and GT.

FUNDING

This research was funded by the FernUniversität in Hagen.

ACKNOWLEDGMENTS

We thank Yoshie Ito, Takahiro Onimaru, Tadashi Dohi, Masatomo Aiba, Keiko Kawashima, and Hiroshi Arima for their help with recruiting participants for Study 3. We thank Akiko Taguchi for her help with translating materials and coding open-ended responses in Study 3.

REFERENCES

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., and Duan, Y. (2016). Facets of the Fundamental Content Dimensions: Agency with Competence and Assertiveness-Communion with Warmth and Morality. *Front. Psychol.* 7, 1810. doi:10.3389/fpsyg.2016.01810

Abele, A. E., Uchronski, M., Suitner, C., and Wojciszke, B. (2008). Towards an Operationalization of the Fundamental Dimensions of agency and Communion: Trait Content Ratings in Five Countries Considering Valence and Frequency of Word Occurrence. *Eur. J. Soc. Psychol.* 38 (7), 1202–1217. doi:10.1002/ejsp.575

Adachi, T. (2014). Occupational Gender Stereotypes Among university Students: Their Relationships with Self-Efficacy and Gender Role Attitudes. *Jpn. Assoc. Industial/Organizational Psychol. J.* 27 (2), 87–100.

Bedyńska, S., Rycielski, P., and Jabłońska, M. (2021). Measuring Stereotype Threat at Math and Language Arts in Secondary School: Validation of a Questionnaire. *Front. Psychol.* 12, 553964. doi:10.3389/fpsyg.2021.553964

Belanger, A. L., Joshi, M. P., Fuesting, M. A., Weisgram, E. S., Claypool, H. M., and Diekman, A. B. (2020). Putting Belonging in Context: Communal Affordances Signal Belonging in STEM. *Pers Soc. Psychol. Bull.* 46 (8), 1186–1204. doi:10.1177/0146167219897181

Bosson, J. K., Jurek, P., Vandello, J. A., Kosakowska-Berezecka, N., Olech, M., Besta, T., et al. (2021). Psychometric Properties and Correlates of Precarious Manhood Beliefs in 62 Nations. *J. Cross-Cultural Psychol.* 52 (3), 231–258. doi:10.1177/0022022121997997

Castaño, A. M., Lubiano, M. A., and García-Izquierdo, A. L. (2020). Gendered Beliefs in STEM Undergraduates: A Comparative Analysis of Fuzzy Rating versus Likert Scales. *Sustainability* 12 (15), 6227. doi:10.3390/su12156227

Cheryan, S., Master, A., and Meltzoff, A. N. (2015). Cultural Stereotypes as Gatekeepers: Increasing Girls' Interest in Computer Science and Engineering by Diversifying Stereotypes. *Front. Psychol.* 6, 49. doi:10.3389/fpsyg.2015.00049

Cheryan, S., Plaut, V. C., Davies, P. G., and Steele, C. M. (2009). Ambient Belonging: How Stereotypical Cues Impact Gender Participation in Computer Science. *J. Pers Soc. Psychol.* 97 (6), 1045–1060. doi:10.1037/a0016239

Cheryan, S., Ziegler, S. A., Montoya, A. K., and Jiang, L. (2017). Why Are Some STEM fields More Gender Balanced Than Others? *Psychol. Bull.* 143 (1), 1–35. doi:10.1037/bul0000052

Cheryan, S., Plaut, V. C., Handron, C., and Hudson, L. (2013). The Stereotypical Computer Scientist: Gendered media Representations as a Barrier to Inclusion for Women. *Sex Roles* 69 (1-2), 58–71. doi:10.1007/s11199-013-0296-x

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates.

Credit Suisse Research Institute (2019). Global Wealth Report 2019. Retrieved from https://www.credit-suisse.com/about-us/en/reports-research/global-wealth-report.html.

Cundiff, J. L., Vescio, T. K., Loken, E., and Lo, L. (2013). Do gender-science Stereotypes Predict Science Identification and Science Career Aspirations Among Undergraduate Science Majors? *Soc. Psychol. Educ.* 16 (4), 541–554. doi:10.1007/s11218-013-9232-8

Destatis (2019). *Studierende in Mathematik, Informatik, Naturwissenschaft (MINT) und Technik-Fächern*. Wiesbaden, Germany: Students in mathematics, computer science, science and techology (STEM) fields. Retrieved from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Tabellen/studierende-mint-faechern.html.

Diekman, A. B., Steinberg, M., Brown, E. R., Belanger, A. L., and Clark, E. K. (2017). A Goal Congruity Model of Role Entry, Engagement, and Exit: Understanding Communal Goal Processes in STEM Gender Gaps. *Pers Soc. Psychol. Rev.* 21 (2), 142–175. doi:10.1177/1088868316642141

Eagly, A. H., and Karau, S. J. (2002). Role Congruity Theory of Prejudice toward Female Leaders. *Psychol. Rev.* 109 (3), 573–598. doi:10.1037/0033-295x.109.3.573

Eagly, A. H., and Wood, W. (2012). "Social Role Theory," in *Handbook of Theories of Social Psychology, Volume II*. Editors P. van Lange, A. Kruglanski, and E. Higgins (Los Angeles, [Calif.]: SAGE), 458–476. doi:10.4135/9781446249222.n49

Eaton, A. A., Saunders, J. F., Jacobson, R. K., and West, K. (2020). How Gender and Race Stereotypes Impact the Advancement of Scholars in STEM: Professors' Biased Evaluations of Physics and Biology Post-Doctoral Candidates. *Sex Roles* 82 (3-4), 127–141. doi:10.1007/s11199-019-01052-w

Else-Quest, N. M., Hyde, J. S., and Linn, M. C. (2010). Cross-national Patterns of Gender Differences in Mathematics: A Meta-Analysis. *Psychol. Bull.* 136 (1), 103–127. doi:10.1037/a0018053

Gender Equality Bureau Cabinet Office (2017). *Gender Equality White Paper (Summary Version) 2017 Edition*. Tokyo, Japan: White paper on gender equality 2017. Retrieved from http://www.gender.go.jp/about_danjo/whitepaper/h29/gaiyou/html/honpen/b1_s05.html.

Gino, F., Wilmuth, C. A., and Brooks, A. W. (2015). Compared to Men, Women View Professional Advancement as Equally Attainable, but Less Desirable. *Proc. Natl. Acad. Sci. U S A.* 112 (40), 12354–12359. doi:10.1073/pnas.1502567112

Gray, J. A., and McNaughton, N. (2003). *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System*. Oxford: Oxford University Press.

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *J. Pers Soc. Psychol.* 74, 1464–1480. doi:10.1037//0022-3514.74.6.1464

Gudykunst, W. B., and Lee, C. M. (2003). Assessing the Validity of Self Construal Scales. *Hum. Comm Res* 29 (2), 253–274. doi:10.1111/j.1468-2958.2003.tb00838.x

Heilman, M. E., and Caleo, S. (2018). Combatting Gender Discrimination: A Lack of Fit Framework. *Group Process. Intergroup Relations* 21 (5), 725–744. doi:10.1177/1368430218761587

Heilman, M. E. (1983). Sex Bias in Work Settings: The Lack of Fit Model. *Res. Organizational Behav.* 5, 269–298.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). Most People Are Not WEIRD. *Nature* 466 (7302), 29. doi:10.1038/466029a

Ikkatai, Y., Inoue, A., Kano, K., Minamizaki, A., McKay, E., and Yokoyama, H. M. (2019). Parental Egalitarian Attitudes towards Gender Roles Affect Agreement on Girls Taking STEM fields at university in Japan. *Int. J. Sci. Edu.* 41 (16), 2254–2270. doi:10.1080/09500693.2019.1671635

Ikkatai, Y., Minamizaki, A., Kano, K., Inoue, A., McKay, E., and Yokoyama, H. M. (2020). Gender-biased Public Perception of STEM fields, Focusing on the Influence of Egalitarian Attitudes toward Gender Roles. *Jcom* 19 (01), A08. doi:10.22323/2.19010208

Jasko, K., Pyrkosz-Pacyna, J., Czarnek, G., Dukała, K., and Szastok, M. (2020). The STEM Graduate: Immediately after Graduation, Men and Women Already Differ in Job Outcomes, Attributions for success, and Desired Job Characteristics. *J. Soc. Issues* 76 (3), 512–542. doi:10.1111/josi.12392

Kassin, S. M., Fein, S., and Markus, H. (2011). *Social Psychology*. 7th ed. Belmont, Calif., United Kingdom: Wadsworth.

Kemmelmeier, M. (2016). Cultural Differences in Survey Responding: Issues and Insights in the Study of Response Biases. *Int. J. Psychol.* 51 (6), 439–444. doi:10.1002/ijop.12386

Law, F., McGuire, L., Winterbottom, M., and Rutland, A. (2021). Children's Gender Stereotypes in STEM Following a One-Shot Growth Mindset Intervention in a Science Museum. *Front. Psychol.* 12, 641695. doi:10.3389/fpsyg.2021.641695

Lindberg, S. M., Hyde, J. S., Petersen, J. L., and Linn, M. C. (2010). New Trends in Gender and Mathematics Performance: A Meta-Analysis. *Psychol. Bull.* 136 (6), 1123–1135. doi:10.1037/a0021276

Markus, H. R., and Kitayama, S. (1991). Culture and the Self: Implications for Cognition, Emotion, and Motivation. *Psychol. Rev.* 98 (2), 224–253. doi:10.1037/0033-295X.98.2.224

Miller, D. I., Eagly, A. H., and Linn, M. C. (2015). Women's Representation in Science Predicts National Gender-Science Stereotypes: Evidence from 66 Nations. *J. Educ. Psychol.* 107, 631–644. doi:10.1037/edu0000005

Moè, A., Hausmann, M., and Hirnstein, M. (2021). Gender Stereotypes and Incremental Beliefs in STEM and Non-STEM Students in Three Countries: Relationships with Performance in Cognitive Tasks. *Psychol. Res.* 85 (2), 554–567. doi:10.1007/s00426-019-01285-0

Moè, A. (2021). Doubling Mental Rotation Scores in High School Students: Effects of Motivational and Strategic Trainings. *Learn. Instruction* 74, 101461. doi:10.1016/j.learninstruc.2021.101461

Morinaga, Y., Sakata, K., Furukawa, Y., and Fukudome, K. (2017). Mathematics Motivation and Gender Stereotypes of Junior and Senior High School Girls. *Kyoiku shinrigaku kenkyu* 65 (3), 375–387. doi:10.5926/jjep.65.375

Nature Index (2020). The Ten Leading Countries in Natural-Sciences Research. *Nature*. 580, S44.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., et al. (2009). National Differences in Gender-Science Stereotypes Predict National Sex Differences in Science and Math Achievement. *Proc. Natl. Acad. Sci.* 106 (26), 10593–10597. doi:10.1073/pnas.0809921106

OECD (2015). Can the Performance gap between Immigrant and Non-immigrant Students Be Closed? *PISA in Focus* 53, 1–4. doi:10.1787/22260919

Park, J., and Kitayama, S. (2014). Interdependent Selves Show Face-Induced Facilitation of Error Processing: Cultural Neuroscience of Self-Threat. *Soc. Cogn. Affective Neurosci.* 9 (2), 201–208. doi:10.1093/scan/nss125

Rudman, L. A., and Glick, P. (2001). Prescriptive Gender Stereotypes and Backlash toward Agentic Women. *J. Soc. Issues* 57 (4), 743–762. doi:10.1111/0022-4537.00239

Rudman, L. A., and Phelan, J. E. (2008). Backlash Effects for Disconfirming Gender Stereotypes in Organizations. *Res. Organizational Behav.* 28, 61–79. doi:10.1016/j.riob.2008.04.003

Sato, K., Yuki, M., and Norasakkunkit, V. (2014). A Socio-Ecological Approach to Cross-Cultural Differences in the Sensitivity to Social Rejection. *J. Cross-Cultural Psychol.* 45 (10), 1549–1560. doi:10.1177/0022022114544320

Sekaquaptewa, D. (2019). Gender-based Microaggressions in STEM Settings. *NCID Currents* 1 (1), 1. doi:10.3998/currents.17387731.0001.101

Smith, P. B. (2014). "Response Bias(es)," in *Encyclopedia of Quality of Life and Well-Being Research*. Editor A. C. Michalos (Dordrecht: Springer Reference), 5539–5540. doi:10.1007/978-94-007-0753-5_2503

Steffens, M. C., and Jelenec, P. (2011). Separating Implicit Gender Stereotypes Regarding Math and Language: Implicit Ability Stereotypes Are Self-Serving for Boys and Men, but Not for Girls and Women. *Sex Roles* 64 (5-6), 324–335. doi:10.1007/s11199-010-9924-x

Steinmetz, J., Bosak, J., Sczesny, S., and Eagly, A. H. (2014). Social Role Effects on Gender Stereotyping in Germany and Japan. *Asian J. Soc. Psychol.* 17 (1), 52–60. doi:10.1111/ajsp.12044

Stout, J. G., Grunberg, V. A., and Ito, T. A. (2016). Gender Roles and Stereotypes about Science Careers Help Explain Women and Men's Science Pursuits. *Sex Roles* 75 (9-10), 490–499. doi:10.1007/s11199-016-0647-5

Sue, D. W. (2010). *Microaggressions in Everyday Life: Race, Gender, and Sexual Orientation*. Hoboken, N.J.: Wiley.

Thomson, R., Yuki, M., Talhelm, T., Schug, J., Kito, M., Ayanian, A. H., et al. (2018). Relational Mobility Predicts Social Behaviors in 39 Countries and Is Tied to Historical Farming and Threat. *Proc. Natl. Acad. Sci.* 115 (29), 7521–7526. doi:10.1073/pnas.1713191115

U. S. News and World Report LP (2020). Best Countries 2020. Global Rankings, International News and Data Insights. Retrieved from https://www.usnews.com/media/best-countries/overall-rankings-2020.pdf.

Vandello, J. A., and Bosson, J. K. (2013). Hard Won and Easily Lost: A Review and Synthesis of Theory and Research on Precarious Manhood. *Psychol. Men Masculinity* 14 (2), 101–113. doi:10.1037/a0029826

Varnum, M., Grossmann, I., Kitayama, S., and Nisbett, R. E. (2010). The Origin of Cultural Differences in Cognition: Evidence for the Social Orientation Hypothesis. *Curr. Dir. Psychol. Sci.* 19 (1), 9–13. doi:10.1177/0963721409359301

Williams, J. E., and Best, D. L. (1991). *Measuring Sex Steretoypes: A Multination Study*. New York: SAGE.

Yalcinkaya, N. S., and Adams, G. (2020). A Cultural Psychological Model of Cross-National Variation in Gender Gaps in STEM Participation. *Personal. Soc. Psychol. Rev.* 24, 345–370. doi:10.1177/1088868320947005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 19 January 2022
doi: 10.3389/fpsyg.2021.789764

Children's Views About Their Future Career and Family Involvement: Associations With Children's Gender Schemas and Parents' Involvement in Work and Family Roles

Joyce J. Endendijk and Christel M. Portengen*

Child and Adolescent Studies, Utrecht University, Utrecht, Netherlands

OPEN ACCESS

Edited by:
Bernhard Ertl,
Munich University of the Federal Armed Forces, Germany

Reviewed by:
Antonya Gonzalez,
Western Washington University, United States
Bettina Hannover,
Freie Universität Berlin, Germany

****Correspondence:***
Joyce J. Endendijk
J.J.Endendijk@uu.nl

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *05 October 2021*
Accepted: *23 December 2021*
Published: *19 January 2022*

Citation:
Endendijk JJ and Portengen CM (2022) Children's Views About Their Future Career and Family Involvement: Associations With Children's Gender Schemas and Parents' Involvement in Work and Family Roles. Front. Psychol. 12:789764. doi: 10.3389/fpsyg.2021.789764

Substantial gender disparities in career advancement are still apparent, for instance in the gender pay gap, the overrepresentation of women in parttime work, and the underrepresentation of women in managerial positions. Regarding the developmental origins of these gender disparities, the current study examined whether children's views about future career and family involvement were associated with children's own gender schemas (gender stereotypes, gender identity) and parents' career- and family-related gender roles. Participants were 142 Dutch families with a child between the ages of 6 and 12 years old (M = 9.80, SD = 1.48, 60% girls). The families had different compositions (1 parent, 2 parents, 1 to 3 children). Children completed a computer task assessing gender stereotypes about toys and questionnaires on gender identity (i.e., felt similarity to same- and other-gender children) and their views about future career and family involvement. Parents reported their occupation, work hours, and task division in the home, which were combined in a composite variable reflecting gender-typicality of career and family involvement. Generalized estimation equations were used to take into account dependency between family members. Results revealed that parents', and especially mothers', gender-typical career and family involvement was associated with children's gender-typical views about future career and family involvement. In addition, children's felt similarity to the same gender was associated with children's gender-typical expectations about career and family involvement. These findings suggest that parents' career, work hours, and task division in the home, together play an important role in how their children envision their future work and family roles. Children themselves also play an active role in developing this vision for the future by their own gender identity, specifically by how similar they feel to individuals of the same gender. To reduce gender disparities in the occupational and domestic domain, programs need to be designed that focus on parental role modeling in the family as well as children's gender identity development.

Keywords: career, family, aspirations, gender identity, role models, parents, middle childhood

INTRODUCTION

Worldwide, substantial gender disparities in career advancement are still apparent, for instance in the gender pay gap (globally, women get paid approximately 20% less than men; International Labour Organization, 2019a), the overrepresentation of women in parttime work (25% of women compared to 10% of men, OECD, 2019), and the underrepresentation of women in managerial positions (<30%, International Labour Organization, 2019b). In addition, men's involvement in the domestic sphere is clearly lacking behind the involvement of women (Croft et al., 2015). Men's share of unpaid labor in the domestic sphere (i.e., childcare, household tasks) ranges from 22–38% in OECD countries (OECD, 2017b). Gender differences are also evident in the types of occupations men and women hold, with women being underrepresented in STEM fields (i.e., science, technology, engineering, mathematics) and men being underrepresented in fields such as health care and education (OECD, 2017a,b).

More gender equality in occupations, career advancement, and involvement in the domestic sphere is of utmost important for several reasons. First, a balanced engagement in both work and family roles is associated with increased general well-being in men and women (Grzywacz et al., 2008). Second, more gender-diversity in work teams improves team collaboration and performance (Bear and Woolley, 2011). Third, increased involvement of men in the domestic sphere reduces the burden on women, increases relationship satisfaction between partners (Stevens et al., 2001), and positively influences children's cognitive development (Malin et al., 2014).

Early indications of gender differences in involvement with certain types of careers and family can be found in children's gender-typical views about their future (Auger et al., 2005; Croft et al., 2014; Polavieja and Platt, 2014; Block et al., 2018). For instance, boys desired to become an athlete, mechanic, or soldier, whereas girls desired being an actor, hairdresser, or teacher (Polavieja and Platt, 2014). Girls also expected to be more family than career oriented in the future, whereas boys expected to be more career oriented than family oriented (i.e., gender-typical expectations about career and family involvement, Croft et al., 2014; Block et al., 2018). Importantly, longitudinal research shows that childhood career aspirations and expectations are linked over time with the actual attained careers of adults (Trice and McClellan, 1993; Mello, 2008; Lawson et al., 2018).

In order to further our understanding of gender disparities in involvement in career and domestic spheres, the current study examined whether children's views about future career and family involvement were associated with children's own gender schemas and parents' gender-role behavior.

Children's Gender Stereotypes and Identity and Their Views About Future Career and Family Involvement

This research is based on predictions from gender-schema theories about factors in the child itself that might contribute to children's gendered views about their future career and family involvement (Bem, 1981; Martin and Halverson, 1987). According to gender-schema theories children play an active role in their own gender development *via* their gender schemas. Gender schemas are dynamic cognitive structures containing gender-related information that children develop and actively construct based on their own experiences with gender in the social environment. Gender-schema theories also predict that gender schemas provide social standards that guide children's behavior and choices. Therefore, based on gender-schema theories, we broadly expect that children with strong gender schemas hold more gender-typed views about their career and family life. Gender schemas encompass different types of schemas (e.g., gender stereotypes, gender attitudes, gender identity, gender self-concept), but the common element is that they concern how people think about themselves and each other in terms of gender (Tenenbaum and Leaper, 2002). The current study focuses on the following gender schemas: gender stereotypes and gender identity. Our specific expectations for these gender schemas are discussed below.

Children's Gender Stereotypes

First, children's gender stereotypes might play a role in children's views about the future. There is ample evidence that children's gender stereotypes about STEM (science, technology, engineering, mathematics) or intellectual ability are linked to gender-typical educational and career choices and interests (Steffens et al., 2010; Cheryan et al., 2015; Bian et al., 2017; Master et al., 2017). Another, less studied, gender-stereotype domain that is relevant to examine in relation to children's views about their future career and family life is the domain of toys. Strong gender stereotypes about toys have been associated with more gender-typed toy play in children (Weisgram, 2016) and the degree of gender-typed play in preschool has been found to predict adolescents' gender-typed occupational interests 10 years later (Kung, 2021). No studies have been done yet, that directly link children's gender stereotypes about toys to gendered visions of their future selves (Fulcher and Coyle, 2018). However, the congruence principle of gender schema theories assumes that congruence exists between personal gender stereotypes and behaviors (Martin and Dinella, 2012). Based on this principle, one could expect strong gender stereotypes to be associated with children's preference for gender-stereotyped occupations and the development of gender-stereotyped beliefs about children's future career versus family involvement.

Children's Gender Identity

Gender identity is a multidimensional construct, which has recently been conceptualized as involving both a connection to one's own gender as well as to the other gender (Martin et al., 2017). The dual-identity conceptualization has been found to be particularly useful for describing individual differences in the (relative) extent to which children report to feel similar to peers of their own biological gender group and to peers of the opposite binary gender group. While most children feel stronger similarity with peers of their own gender than with peers of the other gender, i.e., report a *gender-typical identity* (Martin et al., 2017), the dual-identity approach also acknowledges experiences of transgender youth who feel

dissimilar to peers of their own biological sex and more similar to peers of the opposite sex (Olson and Gülgöz, 2018). Differences in children's felt (relative) similarity to own and other gender peers have been related to children's social-emotional adjustment, gender-typed behavior, and gender attitudes (Andrews et al., 2016; Martin et al., 2017). A few studies in child and adult samples found that stronger gender identity typicality is associated with more traditional (i.e., gender-typed) occupational interests and career choices (Leaper and Van, 2008; Patterson, 2012; Dinella et al., 2014). An explanation for this congruence between gender identity and behavior/interests, is that children are motivated to make their behavior consistent with the behavior of the group they identify with (Martin and Dinella, 2012). This motivation is fueled by feelings of anxiety and discomfort when one violates the gender stereotypes and roles associated with one's gender identity (Akerlof and Kranton, 2000). Further research is needed into how a dual-identity conceptualization of gender identity is associated with children's views about their future career as well as family life.

Parents as Models for Gender Roles

Social learning theories (Bandura, 1977; Bussey and Bandura, 1999) stress the importance of the social context in gender development. Central to these theories is the concept of observational learning and imitation of available models (especially same-gender models) in the child's environment. In the family context, parents are models for future adult gender roles, for example, through the occupations they hold, the hours they work outside the home, as well as how they divide the tasks in the domestic sphere. By observing and imitating the differences between mothers and fathers in involvement with career and family, children will learn how males and females act which will shape their views about future career and family life. In addition, the modeling effect is supposed to be most likely for same-gender models because same-gender models provide information about what are appropriate behaviors for one's own gender (Bandura, 1977). Children might therefore be more likely to model and internalize the occupations, work hours, and task division of same-gender parents in their anticipated career and family orientation.

Previous research indeed showed that different aspects of parents own career and family involvement (i.e., gender-typicality of occupation, work hours, task division at home) are associated with children's views about future career and family involvement (Fulcher and Coyle, 2011; Croft et al., 2014; Polavieja and Platt, 2014; Platt and Polavieja, 2016; Oliveira et al., 2020). In general, these studies showed that more gender-typical engagement of parents with work and the family was associated with more gender-typical views about future career and family life in children. Yet, some studies did not find evidence for these associations (Fulcher, 2011; Croft et al., 2014). In addition, some studies demonstrated that the association between parents' career and family involvement and children's views about future career and family involvement was most salient in same-gender parent-child dyads (Fulcher and Coyle, 2011; Polavieja and Platt, 2014; Oliveira et al., 2020). Yet, there are also studies that did not find evidence for this same-gender modeling effect (Fulcher, 2011; Croft et al., 2014; Platt and Polavieja, 2016).

A possible explanation for the inconsistencies in previous research is that these studies did not examine the family as a system. Instead, in one study mother-child and father-child dyads were analyzed separately and siblings were treated as independent subjects (Croft et al., 2014). Other studies did not take into account dependency between mother and father in a family and/or focused on only one child in each family (Fulcher, 2011; Fulcher and Coyle, 2011; Polavieja and Platt, 2014; Oliveira et al., 2020). Such approaches fail to capture the richness of the family unit, reduce the overall statistical power as not all family members are included in a single analysis, and do not take into account dependency between family members. These issues can confound the effects that were found in previous research. The current study will therefore examine associations between children's gendered views about future career and family involvement, children's gender stereotypes and identity, and parents' gender-role behaviors in the family as a whole.

Middle Childhood as an Important Period for Studying Correlates of Children's Views About Future Career and Family Involvement

Middle childhood (usually defined as ages 6—12) is an important, yet understudied, period for children's gender development (Schroeder and Bámaca-Colbert, 2019) and specifically for children's gendered views about the future. In middle childhood, there are some indications that parental factors appear to be stronger predictors of career aspirations than children's own gender schemas (Croft et al., 2014) or personal attributes (Polavieja and Platt, 2014). Importantly, gender-typed views about future career and family life seem to become more gender-neutral toward the end of middle childhood and into adolescence, particularly for girls (Sandberg et al., 1991; Auger et al., 2005). In addition, in middle childhood children begin to develop gendered self-concepts, with boys seeing themselves as less communal and more agentic than girls (Block et al., 2018). The importance of gender as part of the self-concept appears to increase into early adolescence and reduces into later adolescence (Montemayor and Eisen, 1977). These gendered self-concepts could explain gender differences in anticipated prioritization of family over career in the future (Block et al., 2018), enrollment in male- and female-dominated high-school programs (Tellhed et al., 2018), and choices for STEM careers (Eccles and Wang, 2016). Moreover, children's gender stereotypes increase between age 3 to 5 (Halim et al., 2013), peak between age 5 to 7, and become more flexible during middle childhood (Trautner et al., 2005) and flexibility continues to develop into adolescence (Bartini, 2006). Finally, after being able to identify one's own gender around age 3, and an understanding of gender constancy at 6–7 years of age, in middle childhood children develop a more complex and multidimensional gender identity (Halim and Ruble, 2010). All these developments make middle childhood an appealing setting for studying predictors of children's gendered views about their future life.

Current Study: Research Questions and Hypotheses

In sum, the current study employed a family-systems approach to examine child and family correlates of children's (6–12-year-old) views about their future career and family involvement. Correlates at the child level consisted of gender stereotypes about toys and gender identity. Correlates at the family level consisted of the gender-typicality of mothers' and fathers' occupation (i.e., the proportion of same-gender individuals that work in a certain occupational domain), work hours (i.e., mothers working parttime, fathers working full-time), and task division at home (i.e., degree to which mothers are more responsible for household and child-care tasks than fathers). The following hypotheses were tested:

(1) Children with strong gender schemas (i.e., traditional gender stereotypes about toys, gender-typical identity) hold more gender-typical views about one's future career and family involvement.
(2) Gender-typicality of parents' occupations, work hours, and task division at home is associated with more gender-typical views about future career and family involvement of children.
(3) Associations between gender-typicality of parents' occupations, work hours, and task division at home on the one hand, and children's gender-typical views about career and family involvement on the other hand, are more salient in same-gender dyads than in mixed-gender dyads.

MATERIALS AND METHODS

Participants

Student assistants (BA and MA students in Clinical, Child, Family, and Education studies at Utrecht University) used their personal networks to recruit Dutch families with at least one child between the ages of 6 and 12 years old for this study. Families were contacted *via* information letters (provided in-person or *via* e-mail). The student assistants recruited 142 families. Recruitment and data collection took place between September 2018 and June 2021. The only exclusion criterion was not being able to understand or read Dutch instructions.

From each participating family, one parent (n = 36) or two parents (n = 106) participated. In total, 139 mothers and 108 fathers participated. Regarding the number of participating children per family, in about half of the families (55%, n = 78) only one child was between the ages of 6–12. In 42% (n = 60) of families two children were in the target age range, and in 3% (n = 4) of families three children were in the target age range. **Table 1** presents the background characteristics of this sample. Generally, the majority of the parents in the sample were highly educated.

Procedure

Families were visited at their home by the student assistant who recruited the family. Participants provided written informed consent for their participation at the beginning of the home visit. Each family member subsequently completed questionnaires and a computer task (see section "Instruments") *via* LimeSurvey on a laptop or desktop (duration: approximately 15 min). Parents completed the questionnaires and computer task independently by following the instructions that were presented to them in the LimeSurvey environment. Children completed the questionnaires and computer task under supervision of the student assistant who gave the child verbal instructions. Families received no compensation for their participation. The Ethics Committee of the Faculty of Social Sciences at Utrecht University approved the study (number FETC18-097).

Instruments

Children's Views About Future Career and Family Involvement

Two aspects of children's views about one's future career and family involvement were measured. First, to assess the gender-typicality of children's desired future career, children were asked the following question: "What do you want to be when you grow up?" Children's free responses to this question were coded for gender-typicality of the desired career/occupation. Therefore, we used Dutch Central Bureau of Statistics (CBS), 2021 data that provides information on the proportion of men and women in

TABLE 1 | Sample characteristics.

Family characteristics	
Number of children, range (*M*)	1–5 (2.35)
Gender composition of children, *n* (%)	
All girls	26 (18)
All boys	33 (23)
Mixed gender composition	83 (59)
Family composition, *n* (%)	
Heterosexual two-parent family	127 (90)
Single parent or divorced	15 (10)
Child characteristics	
Age, *M* (*SD*)	9.80 (1.48)
Female gender, *n* (%)	125 (60)
Mothers' characteristics	
Age, *M* (*SD*)	42.44 (4.92)
Educational level, *n* (%)[a]	
Primary education	1 (1)
Lower secondary education	10 (7)
Higher secondary education	40 (29)
Higher vocational education	47 (34)
University	41 (29)
Fathers' characteristics	
Age, *M* (*SD*)	44.42 (5.16)
Educational level, *n* (%)[a]	
Primary education	–
Lower secondary education	8 (7)
Higher secondary education	30 (28)
Higher vocational education	38 (35)
University	32 (30)

[a]*Educational levels are sorted from lowest to highest level.*

an extensive list of occupations. The proportion of women or men (depending on the child's gender) in a certain occupation that corresponded with the occupation mentioned by the child was used for our analyses. Higher scores (>0.50) indicated more gender-typicality of a certain occupation, lower score (<0.50) represent more gender-atypical occupations. When children indicated multiple occupations, the proportions were averaged. In case children answered the question with "I don't know" (or something similar) a proportion of 0.5 (i.e., neutral score) was used for this child.

Second, to assess children's expectations about relative future involvement with career versus family, children were presented with two own-gender individuals and a description of their career and family life (Croft et al., 2014, see **Supplementary Figure 1** for an example). For each pair of individuals (i.e., two pairs were used), a person who worked full time was contrasted with a person who stayed at home caring for the children. Children were asked to indicate for each pair of individuals who they think they will be more like when they are grown up. They rated their similarity on a 5-point scale (1 = most similar to career-oriented target, 2 = a bit more similar to the career-oriented target, 3 = equally similar to both targets, 4 = a bit more similar to the family-oriented target, 5 = most similar to the family-oriented target). Scores were recoded separately for boys and girls and averaged over the two items, in such a way that higher average scores indicated gender-typical expectations of family versus career involvement (i.e., for boys more involvement with career than with family, for girls more involvement with family than with career).

Child Gender Stereotypes About Toys

Children completed a computer task (action inference paradigm; Endendijk et al., 2013) to assess gender stereotypes about toys. The validity of this task to assess gender stereotypes about toys in parents and children has been demonstrated (Endendijk et al., 2013). Participants were asked to divide toys (see **Supplementary Table 1** for a list of toys used) between two fictitious children as quickly as possible, by means of pressing one of two keys on the keyboard ("e" or "i") that were assigned to each child. Pictures of the two children (full color) were presented constantly in the left- and right-hand upper corners of the computer screen. Each full-color toy was presented in the middle of the screen until the participant hit the response key, after which the next full-color toy emerged on the screen.

The task started with a practice block (20 trials) in which red and blue presents had to be divided between two gender-neutral children (could be labeled as both a boy or a girl), followed by two stereotype-congruent blocks and two stereotype-incongruent blocks (17 trials in each block). In the congruent blocks, participants were instructed to assign stereotypically feminine toys (e.g., doll) to a girl and stereotypically masculine toys (e.g., car) to a boy. In the incongruent blocks, participants were instructed to assign stereotypically feminine toys to a boy and stereotypically masculine toys to a girl. To reduce order effects of the presentation of congruent and incongruent blocks (Nosek et al., 2005), the two congruent blocks alternated with the two incongruent blocks (i.e., congruent-incongruent-congruent-incongruent) so that participants made each possible switch between congruent and incongruent blocks. Participants were given a rest period between each block of self-determined length (instructions for the next block were provided in this rest period as well).

The Action Inference Paradigm (AIP) is similar in design to the widely used Implicit Association Test (IAT), as in both tasks a prepotent response tendency (e.g., sort stimuli in a stereotype-consistent way) may either facilitate or interfere with the response required in the task, which in turn influences the speed and accuracy of participants' responses. The main difference between the IAT and AIP is that in the AIP participants must sort only one type of stimuli (e.g., toys) between two categories, whereas in IATs two types of stimuli (e.g., concepts, such as male and female names, and attributes, such as career and family words) must be sorted between two categories, which might be more difficult for children.

The improved scoring algorithm of Greenwald et al. (2003) was used to determine the level of gender stereotypes of the participant. More details about the scoring can be found in the **Supplementary Material**. In short, the gender stereotype score calculated with this algorithm reflects the difference in response latencies between stereotype-incongruent blocks and stereotype-congruent blocks (divided by the pooled SD of response latencies across all trials). Higher scores indicate stronger stereotypical ideas about the appropriateness of certain toys for girls and boys.

Child Gender Identity

Children completed a dual gender identity questionnaire developed and validated by Martin et al. (2017). Participants answered 10 questions regarding how similar they felt to both boys and girls (e.g., "How similar do you feel to [boys/girls]?") using a graphical response scale with circles indicating the level of similarity. Participants answered questions about similarity in five domains: general similarity, behavior, appearance, activities, spending time together. Responses ranged from 0 (circles farthest apart) to 4 (overlapping circles). Separate composite scores were created for the 5 items reflecting similarity to the same-gender group and for the 5 items reflecting similarity to the other-gender group. Higher average scores on these scales reflect more similarity. Reliability of the two scales was good (Cronbach's alpha = 0.85, 0.82, for respectively same-gender and other-gender similarity).

Gender-Typicality of Parents Career and Family Involvement

Parents' Occupation

Parents were asked to report their current occupation. Their responses to this question were coded for gender-typicality the same way as we coded children's aspired occupations, by using CBS data. Twenty percent (n = 100) of the total number of reported careers by parents and children were double coded independently by the first and second authors. The intraclass correlation coefficient (ICC = 0.90) demonstrated excellent coder reliability across this subset of careers. The first author coded the remainder of the careers.

Parents' Work Hours

Parents were asked to report their working hours (i.e., for paid work) per week. Mothers' work hours were inversed (maximum work hours of mothers in this sample subtracted from each mother's work hours) so that higher scores represented more traditional work behaviors (i.e., working less hours outside the house).

Parents' Task Division in the Home

Parents filled out a 15-item questionnaire on their perception of the division of labor regarding small household tasks (e.g., buying groceries, cooking dinner, cleaning) and child-care tasks (e.g., bring children to bed, bathe children, bring children to school) during the past week (Endendijk et al., 2018). Parents could answer on a five-point scale (1 = I exclusively/almost exclusively performed this task, 5 = my partner exclusively/almost exclusively performed this task). Scores on the 15 items were recoded and averaged in such a way that mean scores around 3 represent an egalitarian task division, scores above 3 represent more maternal involvement in the family and scores below 3 represent more paternal involvement in the family. Reliability of this scale in the current study was good (Cronbach's alpha = 0.86). Single parents (n = 6) were also asked to complete this questionnaire. There was variation in their mean scores (range = 1.67–5.00). We checked whether exclusion of these families influenced our results, but this was not the case. Therefore, we decided to keep these families in our sample.

Creation of Composite Gender-Typicality Variable

Following Fulcher and Coyle (2011), we combined gender-typicality of parents' occupation, work hours, and task division in one aggregate variable. First, scores on each variable were recoded in such a way that higher scores reflected more gender-typicality (e.g., more gender-typical occupation, higher paternal work hours, lower maternal work hours, more maternal involvement with small-household and child-care tasks). Second, recoded scores for each variable were standardized into Z-scores and subsequently averaged to create a variable reflecting gender-typicality of parents' career and family involvement. This approach would reduce the number of predictors entered in further analyses. Results of analyses on separate career (work hours, occupation) and family (task division) variables are presented in the Supplementary Material (**Supplementary Tables 2**, **3**, effects are in the same direction but are no longer significant, p-values between 0.105 and 0.173).

Analyses

Generalized Estimating Equations (GEE) were used to analyze the data (Homish et al., 2010) in SPSS (version 24). GEE models are regression-based models that can take into account dependency between variables, such as in family data (Homish et al., 2010). GEE models are more flexible for missing data compared to other models (Zeger et al., 1988) and are therefore suitable for our family data with different family compositions. In addition, in case of a small number of observations in each cluster (i.e., sparse data) GEE is a more robust alternative to multilevel modeling (McNeish, 2014). The family data in the current study could be considered as sparse data because the number of observations per family (i.e., family members) ranged from two to five. Other advantages of GEE over multilevel models include easier model computation and interpretation, more robustness to model misspecification, and no need to model random effects that are not of interest for the research question (McNeish et al., 2017). GEE has been applied to analyze family data in samples ranging from as small as 47 families (Abraham et al., 2021) up to 191 families (Rossen et al., 2018).

Two separate GEEs were conducted, one for children's desired future career and one for children's expectations about future career and family involvement. Each model included main effects for children's gender stereotypes and gender identity (similarity to same- and other gender), and gender-typicality of parents' career and family involvement. In addition, we added a two-way interaction between parent gender and gender-typicality of parents' career and family involvement. This allowed for testing whether associations between gender-typicality of parents' career and family involvement and children's views about future career and family involvement were driven primarily by mothers or fathers. Finally, we added a two-way interaction between gender-composition of the parent-child dyad (same-gender vs. mixed-gender) and gender-typicality of parents' career and family involvement. This enabled testing whether the associations between gender-typicality of parents' career and family involvement and children's views about future career and family involvement were stronger for same-gender dyads. The GEE models were specified with a Gaussian distribution with an identity link for each family, as the dependent variables were continuous (Homish et al., 2010). An exchangeable correlation structure was considered to be most appropriate for the family data (Homish et al., 2010; McNeish et al., 2017). Robust standard errors (Hubert/White Sandwich Estimators) were computed to ensure valid estimations even in case of a mis-specified correlation structure. Parameter estimates were presented as regression coefficients, so that the analyses could be interpreted the same as general linear regression models. For each analysis, we determined which covariates needed to be included based on the change-in-estimate method, >5% change criterion (Rothman et al., 2008).

RESULTS

Descriptive Statistics

Tables 2, **3** display descriptive statistics and correlations for all study variables, for children and parents separately. All variables approached a normal distribution. Several outliers were identified (gender stereotypes: n = 2, other-gender similarity: n = 2, task division: n = 2). These outliers were winsorized (highest non-outlying number + difference between highest non-outlying number and before highest non-outlying number; Tabachnick and Fidell, 2012).

As can be seen in **Table 2**, for children, same-gender similarity was significantly associated with more gender-typical expectations about future career versus family involvement.

TABLE 2 | Descriptive statistics of child study variables.

	1.	2.	3.	4.	*M* (*SD*)
1. Gender-typical desired future career					0.59 (0.20)
2. Gender-typical expectation about career-family	0.10				3.11 (0.84)
3. Gender stereotypes about toys	0.00	−0.08			0.22 (0.33)
4. Same-gender similarity	0.07	0.19**	0.04		4.00 (0.88)
5. Other-gender similarity	−0.07	0.03	−0.15*	−0.54**	2.11 (0.79)

*$p < 0.05$; **$p < 0.01$.

TABLE 3 | Descriptive statistics of parent study variables.

				Mothers	Fathers
	1.	2.	3.	*M* (*SD*)	*M* (*SD*)
1. Gender-typical career		−0.23**	0.09	0.63 (0.20)	0.68 (0.23)
2. Work hours	0.36**		−0.32**	25.81 (9.81)	38.51 (10.16)
3. Gender-typical task division	0.24*	−0.29**		3.79 (0.58)	3.51 (0.64)

Correlations above the diagonal are for mothers. Correlations below the diagonal are for fathers.
*$p < 0.05$; **$p < 0.01$.

Same-gender similarity was negatively associated with other-gender similarity. More other-gender similarity was associated with less strong gender stereotypes about toys. None of the other child variables were significantly correlated. Independent *t*-tests were conducted to examine gender differences on the child variables. First, the proportion of women in girls' desired careers ($M = 0.59$, $SD = 0.20$) was significantly higher than the proportion of women in boys' desired careers [$M = 0.41$, $SD = 0.20$, $t(208) = 6.54$, $p < 0.001$]. Boys were thus more likely to desire careers in which men were overrepresented, whereas girls were more likely to desire careers in which women were overrepresented. Second, girls expected more gender-typical family versus career involvement ($M = 3.29$, $SD = 0.82$) than boys [$M = 2.85$, $SD = 0.79$, $t(208) = 3.85$, $p < 0.001$]. Girls also reported more other-gender similarity ($M = 2.28$, $SD = 0.76$) than boys did [$M = 1.88$, $SD = 0.77$, $t(208) = 3.70$, $p < 0.001$]. However, boys reported more same-gender similarity ($M = 4.16$, $SD = 0.76$) than girls did [$M = 3.88$, $SD = 0.94$, $t(201.04) = -2.38$, $p = 0.018$]. There was no gender difference in children's gender stereotypes [$t(208) = -0.81$, $p = 0.418$] (i.e., boys and girls did not differ in response latencies to stereotype-inconsistent versus stereotype-consistent trials in the task assessing gender stereotypes about toys).

As can be seen in **Table 3**, for fathers, there were significant associations in the expected direction between work hours, task division, and gender-typicality of their career. For mothers, more work hours were associated with a less traditional task division as well as a less gender-typical career. An independent *t*-test on the proportion of women in the occupations that fathers and mothers reported themselves to be in, revealed that mothers reported occupations with a higher proportion of women ($M = 0.63$, $SD = 0.20$) than fathers [$M = 0.32$, $SD = 0.23$, $t(208,76) = 11.28$, $p < 0.001$].

As all the correlations between the independent variables in **Tables 2**, **3** were below 0.70, there were no issues with multi-collinearity in further analyses.

Predictors of the Gender-Typicality of Children's Desired Future Career

Table 4 displays results for the final GEE model for children's gender-typical desired future career. Only parents' gender-typical career and family involvement was associated with the gender-typicality of children's desired career. Children's gender stereotypes about toys and same- and other-gender similarity were not related to children's desired career. Regarding the covariates, younger child age and older parental age were associated with more gender-typical desired careers.

The additional interaction between parent gender and parents' gender-typical career-family involvement was not significant

TABLE 4 | Generalized estimation equations predicting gender-typicality of children's desired career from children's gender identity, stereotypes, and parents' gender-typical career and family involvement.

	B	*SE*	95% CI	Wald	*p*
Child gender[1]	0.01	0.04	[−0.06, 0.08]	0.11	0.741
Child age	−0.03*	0.01	[−0.05, −0.004]	5.65	0.017
Parent gender[2]	0.01	0.01	[−0.003, 0.02]	2.11	0.146
Parent age	0.004*	0.002	[0.00, 0.01]	4.09	0.043
Educational level[3]					
Primary education	−0.003	0.04	[−0.08, 0.08]	0.01	0.934
Lower secondary education	0.02	0.05	[−0.08, 0.13]	0.18	0.676
Higher secondary education	−0.06	0.04	[−0.13, 0.01]	2.68	0.101
Higher vocational education	−0.02	0.04	[−0.09, 0.06]	0.19	0.661
Family composition[4]					
Single parent/divorced	0.05	0.03	[−0.01, 0.12]	2.38	0.123
Child gender stereotypes about toys	−0.01	0.04	[−0.10, 0.08]	0.05	0.818
Child same-gender similarity	0.01	0.02	[−0.04, 0.06]	0.17	0.679
Child other-gender similarity	−0.01	0.02	[−0.05, 0.03]	0.19	0.663
Gender-typicality of parents' career and family involvement[5]	0.03*	0.01	[0.003, 0.05]	4.94	0.026

[1] *Boys are reference category.*
[2] *Fathers are reference category.*
[3] *University level was the reference category.*
[4] *Two-parent family was the reference category.*
[5] *This variable is a standardized composite score including gender-typicality of work hours, gender-typicality of occupation, and gender-typicality of task division.*
*$p < 0.05$.

TABLE 5 | Generalized estimation equations predicting children's gender-typical expectations about future career and family involvement from children's gender identity, stereotypes and parents' gender-typical career and family involvement.

	B	*SE*	95% *CI*	Wald	*p*
Child gender[1]	0.38*	0.14	[0.12, 0.65]	7.93	0.005
Child age	−0.11*	0.04	[−0.18, −0.04]	8.49	0.004
Parent gender[2]	−0.03	0.03	[−0.09, 0.03]	1.03	0.310
Parent age	−0.01	0.01	[−0.03, 0.01]	0.50	0.482
Family gender composition[3]					
All boys	−0.08	0.17	[−0.41, 0.25]	0.22	0.641
All girls	0.03	0.15	[−0.26, 0.33]	0.47	0.828
Child gender stereotypes about toys	−0.27	0.16	[−0.59, 0.04]	2.92	0.087
Child same-gender similarity	0.24*	0.08	[0.07, 0.40]	8.17	0.004
Child other-gender similarity	0.11	0.08	[−0.05, 0.28]	1.78	0.182
Gender-typicality of parents' career and family involvement	−0.09	0.06	[−0.21, 0.03]	2.12	0.146
Parent gender*Gender-typicality career-family involvement[2]	0.10*	0.05	[0.01, 0.19]	4.49	0.034

[1]*Boys are reference category.*
[2]*Fathers are reference category. This variable is a standardized composite score including gender-typicality of work hours, gender-typicality of occupation, and gender-typicality of task division.*
[3]*Mixed gender composition of children is the reference category.*
*$p < 0.05$.

($B = -0.001$, $SE = 0.01$, 95% CI = −0.02, 0.02, Wald = 0.002, $p = 0.962$). This indicated that the association between parents' gender-typical career-family involvement and children's desired future career was not driven primarily by mothers or fathers. The additional interaction between gender composition of the parent-child dyad and parents' gender-typical career-family involvement was not significant ($B = -0.01$, $SE = 0.02$, 95% CI = −0.04, 0.02, Wald = 0.15, $p = 0.704$). This indicated that the association between parents' gender-typical career-family involvement and children's desired career was not different for same-gender and other-gender parent-child dyads.

Predictors of the Gender-Typicality of Children's Expectations About Future Career and Family Involvement

Table 5 displays results for the final GEE model for children's gender-typical expectations about involvement with career and family. More same-gender similarity in children was associated with more gender-typical expectations about involvement with career and family. Children's gender stereotypes about toys and other-gender similarity were not related to children's gender-typical career-family expectations. Regarding the covariates, being a girl and younger child age were associated with more gender-typical expectations about future career and family involvement.

In addition, the interaction between parent gender and parents' gender-typical career-family involvement was significant ($B = 0.10$, $SE = 0.05$, 95% CI = 0.01, 0.19, Wald = 4.49, $p = 0.034$). This indicated that only mothers' gender-typical career and family involvement was associated with children's gender-typical expectations about future career and family involvement.

The additional interaction between gender composition of the parent-child dyad and parents' gender-typical career-family involvement was not significant ($B = -0.05$, $SE = 0.07$, 95% CI = −0.19, 0.10, Wald = 0.38, $p = 0.536$). This indicated that the association between parents' gender-typical career-family involvement and children's expected career-family involvement was not different for same-gender and mixed-gender parent-child dyads.

DISCUSSION

This study was conducted to examine whether children's views about future career and family involvement were associated with children's own gender stereotypes and identity as well as parents' gender-typical career and family involvement. Results revealed that parents', and especially mothers', gender-typical career and family involvement was associated with children's gender-typical views about their future career and family life. In addition, children's felt similarity to the same gender as well as mothers' gender-typical career and family involvement were associated with children's gender-typical expectations about their future career and family involvement. Children's gender stereotypes about toys were not related to children's views about future career and family involvement. Finally, associations between parent's gender-typical career and family involvement and children's views about their future were not different between same-gender and mixed-gender parent-child dyads.

Our findings for children's gender identity provide some support for gender-schema theories' prediction that gender schemas provide social standards that guide children's behavior and choices (Bem, 1981; Martin and Halverson, 1987). Children with strong gender schemas, for example because they felt high similarity to same-gender peers, in this study indeed held more gender-typical expectations about future career and family involvement but did not desire a more gender-typical career. An explanation for the congruence between the level of same-gender similarity and children's gender-typical expectations about their career and family involvement, is that children are motivated to make their behavior consistent with the behavior of the group they identify with (Martin and Dinella, 2012). That desired career was not linked with children's gender identity might be because children in middle childhood children still have limited knowledge of the gender typicality of occupations (Gottfredson, 2002). Our findings extend previous research linking higher gender-typicality to more traditional occupational interests and career choices (Leaper and Van, 2008; Patterson, 2012; Dinella et al., 2014), by showing that gender identity aspects also relate to expectations about future involvement in the domestic sphere.

Unexpectedly, children's gender stereotypes, specifically in relation to toys, were not related to their views about their future career and family involvement. It could be that the link between children's gender stereotypes about toys and children's views about their future is too indirect to be found without also examining possible underlying mediating factors.

For instance, gender stereotypes about toys have been associated with gender-typed toy play (Weisgram, 2016) which in turn has been associated with adolescents' gender-typical occupational interests (Kung, 2021). Future research could examine this mediational process. The lack of associations with children's gender stereotypes about toys might also be due to our measure including both toys that have a clear link with the domestic sphere (e.g., baby dolls, toy kitchen) or the career sphere (e.g., fire truck, tools), as well as toys that are less directly linked to these domains (e.g., pirate costume, princess costume). Our measure consisted of too few trials to examine the effect of toy type. Future research could examine whether children's gender stereotypes about toys with clear links to the career or domestic spheres are related to their views about future career and family involvement.

We also found some evidence for the role modeling prediction from social learning theory (Bandura, 1977; Bussey and Bandura, 1999). It appears that parents', and especially mothers', gender-typical career and family involvement are associated with children's views about future career and family life. Our findings demonstrate that previously found associations between parents' work- and family-related gender roles and children's career and family aspirations (Fulcher and Coyle, 2011; Croft et al., 2014; Polavieja and Platt, 2014; Oliveira et al., 2020) also hold in a family-systems context. In the current study parents could provide a model for traditional gender-role behavior by working in a career domain with a high percentage of same-gender peers, when mothers worked few hours outside the home, when fathers worked many hours outside the home, and when mothers were more responsible than fathers for household and childcare tasks. By observing such traditional gender roles in the career *and* family involvement of their parents, children will learn how males and females act, which will shape their views about their future career and family life. An explanation for why especially mothers' work- and family-related gender roles were important for children's expectations about career versus family involvement could be that especially mothers might provide a model for balancing work and family roles. Indeed, mothers have been found to experience more work-family conflict than fathers (Shockley et al., 2017).

No support was found for the same-gender modeling hypothesis of social learning theory (Bandura, 1977; Bussey and Bandura, 1999). Previous research also produced mixed findings regarding same-gender modeling of parents' career and family involvement (Fulcher, 2011; Fulcher and Coyle, 2011; Croft et al., 2014; Polavieja and Platt, 2014; Oliveira et al., 2020). In the current study the associations between parents' gender-typical career and family involvement and children's views about future career and family involvement, were not more salient in same-gender dyads than in mixed-gender dyads. It appears that fathers and mothers are important role models for both boys and girls. This might not be surprising as mothers' and fathers' gender roles in a family are closely interrelated (Oláh and Neyer, 2021). For instance, when one parent increases their work hours, the other parent is likely to compensate for the reduced involvement in the family (Hook, 2006; Fox, 2009). So, it might actually be the combination of mothers and fathers work in and outside the family that conveys messages to children about how men and women balance work and family responsibilities and that shapes children's views about future career and family life.

A final noteworthy finding is that we found correlational evidence for a possible developmental process implicated in children's views about their future selves, as older child age was associated with less gender-typical views about career and family. This finding fits with previous research demonstrating that children's gender stereotypes become more flexible and less rigid over time (Trautner et al., 2005). In addition, this finding is noteworthy because it could imply that children's views about their future career and family involvement might over time become less congruent with their gender identity or their parents' career and family involvement (assuming that the latter two factors remain relatively constant over time). This hypothesis remains to be tested longitudinally, as well as how children experience or resolve this increasing incongruence.

Even though our study is strong in terms of the family-systems approach and the use of mixed methods (i.e., computer task, parent-report, child-report), our findings must be viewed in light of some limitations. First, because of the correlational design of this study, we were not able to determine the direction of effects in the association that were found. More longitudinal research is now necessary to unravel the gendered developmental processes underlying the career decision making process. Second, our sample size was too small to optimally utilize the dual gender identity approach by examining how different gender-identity typologies are related to children's gender-typical views about career and family involvement. Third, even though we, and previous studies (e.g., Croft et al., 2014), found relevant associations with children's expected future involvement with career and family, the measure used to assess children's expectations only consisted of 2 items. Future research could extend this measure to assess children's gendered expectations for the future in a more multi-faceted way. Finally, a convenience sampling method was used, which resulted in a sample that was more highly educated than the population.

In sum, this family-systems study demonstrated that parents' own career, work hours, and task division in the home, together play an important role in how their children envision their future work and family roles. This suggests that intergenerational transmission plays a role in the perpetuation of gender disparities in the occupational and domestic domain. Children themselves also play an active role in developing this vision for the future by their own gender identity, specifically by how similar they feel to individuals of the same gender. A practical implication of these findings is that parents need to be made aware of the roles their own gender-role behavior, as well as their children's gender identity, play in the career decision making process of their children. For boys and girls to make career decisions that fit with their interests and competencies, instead of their gender or their parents' gender roles, parents could encourage children to explore a wide range of career and educational options. In addition, programs and policies could stimulate more equality in parental gender roles as well as children's felt similarity to people of both genders, in order to reduce gender disparities in the occupational and domestic domain.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of the Faculty of Social Sciences at Utrecht University (number FETC18-097). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

JE: conceptualization, formal analysis, methodology and design, and writing – original draft. JE and CP: supervision of data collection and processing of data. CP: writing – review and editing. Both authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.789764/full#supplementary-material

REFERENCES

Abraham, E., Zagoory-Sharon, O., and Feldman, R. (2021). Early maternal and paternal caregiving moderates the links between preschoolers' reactivity and regulation and maturation of the HPA-immune axis. *Dev. Psychobiol.* 63, 1482–1498. doi: 10.1002/dev.22089

Akerlof, G. A., and Kranton, R. E. (2000). Economics and identity. *Q. J. Econ.* 115, 715–753. doi: 10.1162/003355300554881

Andrews, N. C., Martin, C. L., and Gallagher, A. M. (2016). On the association between self-reported own-and other-gender similarity and the use of physical and relational aggression in sixth grade children. *Arch. Sex. Behav.* 45, 1817–1826. doi: 10.1007/s10508-015-0685-z

Auger, R. W., Blackhurst, A. E., and Wahl, K. H. (2005). The development of elementary-aged children's career aspirations and expectations. *Prof. Sch. Couns.* 8, 322–329.

Bandura, A. (1977). *Social Learning Theory*. Englewood Cliffs, NJ: Prentice Hall.

Bartini, M. (2006). Gender role flexibility in early adolescence: developmental change in attitudes, self-perceptions, and behaviors. *Sex Roles* 55, 233–245. doi: 10.1007/s11199-006-9076-1

Bear, J. B., and Woolley, A. W. (2011). The role of gender in team collaboration and performance. *Interdiscip. Sci. Rev.* 36, 146–153. doi: 10.1179/030801811X13013181961473

Bem, S. L. (1981). Gender schema theory: a cognitive account of sex typing. *Psychol. Rev.* 88, 354–364. doi: 10.1037/0033-295X.88.4.354

Bian, L., Leslie, S. J., and Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science* 355, 389–391. doi: 10.1126/science.aah6524

Block, K., Gonzalez, A. M., Schmader, T., and Baron, A. S. (2018). Early gender differences in core values predict anticipated family versus career orientation. *Psychol. Sci.* 29, 1540–1547. doi: 10.1177/0956797618776942

Bussey, K., and Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychol. Rev.* 106, 676–713. doi: 10.1037/0033-295X.106.4.676

Central Bureau of Statistics (CBS) (2021). *Werkzame Beroepsbevolking; Beroep*. Available online at: https://opendata.cbs.nl/#/CBS/nl/dataset/82808NED/table?searchKeywords=beroepsgroep (accessed July 1, 2021).

Cheryan, S., Master, A., and Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: increasing girls' interest in computer science and engineering by diversifying stereotypes. *Front. Psychol.* 6:49. doi: 10.3389/fpsyg.2015.00049

Croft, A., Schmader, T., and Block, K. (2015). An underexamined inequality: cultural and psychological barriers to men's engagement with communal roles. *Pers. Soc. Psychol. Rev.* 19, 343–370. doi: 10.1177/1088868314564789

Croft, A., Schmader, T., Block, K., and Baron, A. S. (2014). The second shift reflected in the second generation: do parents' gender roles at home predict children's aspirations? *Psychol. Sci.* 25, 1418–1428. doi: 10.1177/0956797614533968

Dinella, L. M., Fulcher, M., and Weisgram, E. S. (2014). Sex-typed personality traits and gender identity as predictors of young adults' career interests. *Arch. Sex. Behav.* 43, 493–504. doi: 10.1007/s10508-013-0234-6

Eccles, J. S., and Wang, M. T. (2016). What motivates females and males to pursue careers in mathematics and science? *Int. J. Behav. Dev.* 40, 100–106. doi: 10.1177/0165025415616201

Endendijk, J. J., Derks, B., and Mesman, J. (2018). Does parenthood change implicit gender-role stereotypes and behaviors? *J. Marriage Fam.* 80, 61–79. doi: 10.1111/jomf.12451

Endendijk, J. J., Groeneveld, M. G., van Berkel, S. R., Hallers-Haalboom, E. T., Mesman, J., and Bakermans-Kranenburg, M. J. (2013). Gender stereotypes in the family context: mothers, fathers, and siblings. *Sex Roles* 68, 577–590. doi: 10.1007/s11199-013-0265-4

Fox, B. (2009). *When Couples Become Parents: The Creation of Gender in the Transition to Parenthood*. Toronto, ON: University of Toronto Press.

Fulcher, M. (2011). Individual differences in children's occupational aspirations as a function of parental traditionality. *Sex Roles* 64, 117–131. doi: 10.1007/s11199-010-9854-7

Fulcher, M., and Coyle, E. F. (2011). Breadwinner and caregiver: a cross-sectional analysis of children's and emerging adults' visions of their future family roles. *Br. J. Dev. Psychol.* 29, 330–346. doi: 10.1111/j.2044-835X.2011.02026.x

Fulcher, M., and Coyle, E. F. (2018). "Working at play: gender-typed play and children's visions of future work and family roles," in *Gender Typing of Children's Toys: How Early Play Experiences Impact Development*, eds E. S. Weisgram and L. M. Dinella (Washington, DC: American Psychological Association), 257–286.

Gottfredson, L. S. (2002). "Gottfredson's theory of circumscription, compromise, and self-creation," in *Career Choice and Development*, 4th Edn, ed. D. Brown (San Francisco, CA: Jossey-Bass), 85–148.

Greenwald, A. G., Nosek, B. A., and Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* 85, 197–216. doi: 10.1037/0022-3514.85.2.197

Grzywacz, J. G., Butler, A. B., and Almeida, D. M. (2008). "Work, family, and health: work-family balance as a protective factor against stresses of daily life," in *The Changing Realities of Work and Family: A Multidisciplinary Approach*, eds A. Marcus-Newhall, D. F. Halpern, and S. J. Tan (Hoboken, NJ: Wiley Blackwell), 194–215. doi: 10.1002/9781444305272.ch10

Halim, M. L., and Ruble, D. (2010). "Gender identity and stereotyping in early and middle childhood," in *Handbook of Gender Research in Psychology*, eds J. C. Chrisler and D. R. McCreary (New York, NY: Springer), 495–525.

Halim, M. L., Ruble, D., Tamis-LeMonda, C., and Shrout, P. E. (2013). Rigidity in gender-typed behaviors in early childhood: a longitudinal study of ethnic minority children. *Child Dev.* 84, 1269–1284. doi: 10.1111/cdev.12057

Homish, G. G., Edwards, E. P., Eiden, R. D., and Leonard, K. E. (2010). Analyzing family data: a GEE approach for substance use researchers. *Addict. Behav.* 35, 558–563. doi: 10.1016/j.addbeh.2010.01.002

Hook, J. L. (2006). Care in context: men's unpaid work in 20 countries, 1965-2003. *Am. Sociol. Rev.* 71, 639–660. doi: 10.1177/000312240607100406

International Labour Organization (2019a). *Female Share of Employment in Managerial Positions*. Available online at: https://www.ilo.org/shinyapps/bulkexplorer3/?lang=en&segment=indicator&id=SDG_0552_OCU_RT_A (accessed July 1, 2021)

International Labour Organization (2019b). *Global Wage Report 2018/19: What Lies Behind Gender Pay Gaps*. Available online at: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_650553.pdf (accessed July 1, 2021)

Kung, K. T. (2021). Preschool gender-typed play behavior predicts adolescent gender-typed occupational interests: a 10-year longitudinal study. *Arch. Sex. Behav.* 50, 843–851. doi: 10.1007/s10508-021-01976-z

Lawson, K. M., Lee, B., Crouter, A. C., and McHale, S. M. (2018). Correlates of gendered vocational development from middle childhood to young adulthood. *J. Vocat. Behav.* 107, 209–221. doi: 10.1016/j.jvb.2018.05.002

Leaper, C., and Van, S. R. (2008). Masculinity ideology, covert sexism, and perceived gender typicality in relation to young men's academic motivation and choices in college. *Psychol. Men Masc.* 9, 139–153. doi: 10.1037/1524-9220.9.3.139

Malin, J. L., Cabrera, N. J., and Rowe, M. L. (2014). Low-income minority mothers' and fathers' reading and children's interest: longitudinal contributions to children's receptive vocabulary skills. *Early Child. Res. Q.* 29, 425–432. doi: 10.1016/j.ecresq.2014.04.010

Martin, C. L., and Dinella, L. M. (2012). Congruence between gender stereotypes and activity preference in self-identified tomboys and non-tomboys. *Arch. Sex. Behav.* 41, 599–610. doi: 10.1007/s10508-011-9786-5

Martin, C. L., and Halverson, C. F. (1987). "The roles of cognition in sex role acquisition," in *Current Conceptions of Sex Roles and Sex Typing: Theory and Research*, ed. D. B. Carter (New York, NY: Praeger), 123–137.

Martin, C. L., Andrews, N. C., England, D. E., Zosuls, K., and Ruble, D. N. (2017). A dual identity approach for conceptualizing and measuring children's gender identity. *Child Dev.* 88, 167–182. doi: 10.1111/cdev.12568

Master, A., Cheryan, S., Moscatelli, A., and Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. *J. Exp. Child Psychol.* 160, 92–106. doi: 10.1016/j.jecp.2017.03.013

McNeish, D. M. (2014). Modeling sparsely clustered data: design-based, model-based, and single-level methods. *Psychol. Methods* 19, 552–563. doi: 10.1037/met0000024

McNeish, D., Stapleton, L. M., and Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychol. Methods* 22, 114–140. doi: 10.1037/met0000078

Mello, Z. R. (2008). Gender variation in developmental trajectories of educational and occupational expectations and attainment from adolescence to adulthood. *Dev. Psychol.* 44, 1069–1080. doi: 10.1037/0012-1649.44.4.1069

Montemayor, R., and Eisen, M. (1977). The development of self-conceptions from childhood to adolescence. *Dev. Psychol.* 13, 314–319. doi: 10.1037/0012-1649.13.4.314

Nosek, B. A., Greenwald, A. G., and Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Pers. Soc. Psychol. Bull.* 31, 166–180. doi: 10.1177/0146167204271418

OECD (2017b). *The Pursuit of Gender Equality: An Uphill Battle*. Paris: OECD Publishing.

OECD (2017a). *Education at a Glance 2017: OECD Indicators*. Paris: OECD Publishing.

OECD (2019). *Employment: Share of Employed in Part-Time Employment, By Sex and Age Group*. Paris: OECD Publishing.

Oláh, L. S., and Neyer, G. (2021). "Demographic challenges of Europe in the new millennium: swedish family policies as an answer to them?," in *Challenges to the Welfare State,* eds. J. Aidukaite, S. E. O. Hort, and S. Kuhnle (Edward Elgar Publishing), 33–51.

Oliveira, Í. M., Porfeli, E. J., do Céu Taveira, M., and Lee, B. (2020). Children's career expectations and parents' jobs: intergenerational (dis)continuities. *Career Dev. Q.* 68, 63–77. doi: 10.1002/cdq.12213

Olson, K. R., and Gülgöz, S. (2018). Early findings from the transyouth project: gender development in transgender children. *Child Dev. Perspect.* 12, 93–97. doi: 10.1111/cdep.12268

Patterson, M. M. (2012). Self-perceived gender typicality, gender-typed attributes, and gender stereotype endorsement in elementary-school-aged children. *Sex Roles* 67, 422–434. doi: 10.1007/s11199-012-0184-9

Platt, L., and Polavieja, J. (2016). Saying and doing gender: intergenerational transmission of attitudes towards the sexual division of labour. *Eur. Sociol. Rev.* 32, 820–834. doi: 10.1093/esr/jcw037

Polavieja, J. G., and Platt, L. (2014). Nurse or mechanic? The role of parental socialization and children's personality in the formation of sex-typed occupational aspirations. *Soc. Forces* 93, 31–61. doi: 10.1093/sf/sou051

Rossen, L., Mattick, R. P., Wilson, J., Burns, L., Macdonald, J. A., Olsson, C., et al. (2018). Mother–infant and partner–infant emotional availability at 12 months of age: findings from an Australian longitudinal study. *Infancy* 23, 893–916. doi: 10.1111/infa.12247

Rothman, K. J., Greenland, S., and Lash, T. L. (2008). *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams and Wilkins.

Sandberg, D. E., Ehrhardt, A. A., Ince, S. E., and Meyer-Bahlburg, H. F. (1991). Gender differences in children's and adolescents' career aspirations: a follow-up study. *J. Adolesc. Res.* 6, 371–386. doi: 10.1177/074355489163007

Schroeder, K. M., and Bámaca-Colbert, M. Y. (2019). Cultural underpinnings of gender development: studying gender among children of immigrants. *Child Dev.* 90, 1005–1015. doi: 10.1111/cdev.13265

Shockley, K. M., Shen, W., DeNunzio, M. M., Arvan, M. L., and Knudsen, E. A. (2017). Disentangling the relationship between gender and work–family conflict: an integration of theoretical perspectives using meta-analytic methods. *J. Appl. Psychol.* 102, 1601–1635. doi: 10.1037/apl0000246

Steffens, M. C., Jelenec, P., and Noack, P. (2010). On the leaky math pipeline: comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *J. Educ. Psychol.* 102, 947–963. doi: 10.1037/a0019920

Stevens, D., Kiger, G., and Riley, P. J. (2001). Working hard and hardly working: domestic labor and marital satisfaction among dual-earner couples. *J. Marriage Fam.* 63, 514–526. doi: 10.1111/j.1741-3737.2001.00514.x

Tabachnick, B. G., and Fidell, L. S. (2012). *Using Multivariate Statistics*, 6th Edn. New York, NY: Harper Collins.

Tellhed, U., Bäckström, M., and Björklund, F. (2018). The role of ability beliefs and agentic vs. communal career goals in adolescents' first educational choice. What explains the degree of gender-balance? *J. Vocat. Behav.* 104, 1–13. doi: 10.1016/j.jvb.2017.09.008

Tenenbaum, H. R., and Leaper, C. (2002). Are parents' gender schemas related to their children's gender-related cognitions? A meta-analysis. *Dev. Psychol.* 38, 615–630. doi: 10.1037/0012-1649.38.4.615

Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., and Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: developmental or differential? *Infant Child Dev.* 14, 365–381. doi: 10.1002/icd.399

Trice, A. D., and McClellan, N. (1993). Do children's career aspirations predict adult occupations? An answer from a secondary analysis of a longitudinal study. *Psychol. Rep.* 72, 368–370. doi: 10.2466/pr0.1993.72.2.368

Weisgram, E. S. (2016). The cognitive construction of gender stereotypes: evidence for the dual pathways model of gender differentiation. *Sex Roles* 75, 301–313. doi: 10.1007/s11199-016-0624-z

Zeger, S. L., Liang, K. Y., and Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44, 1049–1060. doi: 10.2307/2531734

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 February 2022
doi: 10.3389/fpsyg.2021.796073

How Parents' Stereotypical Beliefs Relate to Students' Motivation and Career Aspirations in Mathematics and Language Arts

Kathryn Everhart Chaffee and Isabelle Plante*

Département de didactique, Université du Québec à Montréal, Montreal, QC, Canada

Despite progress, gender gaps persist in mathematical and language-related fields, and gender stereotypes likely play a role. The current study examines the relations between parents' gender-related beliefs and their adolescent child's motivation and career aspirations through a survey of 172 parent-child dyads. Parents reported their gendered beliefs about ability in mathematics and language arts, as well as their prescriptive gender role beliefs. Students reported their expectancies and values in these two domains, as well as their career aspirations The results of path models suggested that parents' ability stereotypes about language boosted girls' motivation for language arts, thereby nudging them away from STEM pathways. Girls' career aspirations stemmed not only from their valuation of the corresponding domain, but also from their valuation of competing domains. Such findings highlight the need to consider multiple domains simultaneously in order to better capture the complexity of girls' career decisions. For boys, parents' language ability stereotypes were directly related to mathematical career aspirations. These results suggest that stereotypes that language arts is not for boys push them instead toward mathematics. Our study also highlighted the unique role of parental beliefs in traditional gender roles for boys' motivation and career aspirations. Specifically, parents' gender role stereotypes directly related to less interest in language arts only among boys. This highlights that research into gender gaps in female-dominated fields should consider stereotypes related to appropriate behavior and social roles for boys.

Keywords: gender role, gender stereotype, career interest, parent beliefs, late adolescence, expectancy-value, gender gap

OPEN ACCESS

Edited by:
Daniela Barni,
University of Bergamo, Italy

Reviewed by:
Silvia Di Battista,
University of Bergamo, Italy
Claudia Russo,
Libera Università Maria SS. Assunta, Italy

****Correspondence:***
Kathryn Everhart Chaffee
kathryn.chaffee@ualberta.ca
orcid.org/0000-0001-6381-0228

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *15 October 2021*
Accepted: *21 December 2021*
Published: *04 February 2022*

Citation:
Chaffee KE and Plante I (2022) How Parents' Stereotypical Beliefs Relate to Students' Motivation and Career Aspirations in Mathematics and Language Arts. Front. Psychol. 12:796073. doi: 10.3389/fpsyg.2021.796073

INTRODUCTION

Despite efforts to reduce gender gaps in science, technology, engineering, and mathematics (STEM) fields, women remain underrepresented in STEM careers (Wang et al., 2013; Simon et al., 2016). Contrastingly, men's underrepresentation in female-dominated fields such as those related to language has remained pronounced and stable over time (Croft et al., 2015). These gender imbalances are problematic, as these fields may not adequately benefit from the contributions of the most competent and interested individuals of all genders. Considering the economic and societal

importance of this skewed gender representation, decades of research has focused on understanding the reasons for gendered preferences and aspirations, mostly in STEM domains (e.g., Hyde et al., 1990; Eccles, 1994). Although gender differences in career aspirations are certainly complex and influenced by diverse factors, it is now well-established that social factors play a key role (Hyde, 2014; Olsson and Martiny, 2018; Froehlich et al., 2020).

In accordance with such results, the social cognitive perspective (Bandura, 1977) proposes that gendered interests and aspirations are largely rooted in the social context. In particular, parents, as important socializers, play a crucial role in students' education and development (Šimunović and Babarović, 2020). Their socializing role may also manifest through their attitudes and cultural values, such as those reflected by their gender stereotypical beliefs (Tomasetto et al., 2015). Though such beliefs may be important for shaping students' interests throughout their schooling (Muntoni and Retelsdorf, 2019), they might be particularly important when adolescents need to make decisions about their future and choose between multiple programs. During that time, students may be especially likely to seek the approval and guidance of their parents. Parents' feedback regarding their son's or daughter's career decisions could be colored by the parent's stereotypical beliefs. For example, subtle messages from parents such as "it's tough for women in science" or "you'll be the only boy in your literature program" could have large downstream consequences for students' career decision-making.

The current research seeks to examine the role of parents' stereotypes during the final year of high school, just before students decide to either pursue a stereotypical field (e.g., mathematics for boys) or a counter-stereotypical field (e.g., communication and literature for boys) after graduation. In addition, contrary to most research that relies exclusively on student reports, the current work combines parents' actual self-reported beliefs with students' self-reported motivation and aspirations in the two stereotypical domains of mathematics and language arts. These two domains are particularly relevant to study because they typically receive the greatest curricular emphasis and instructional time throughout mandatory schooling in most Western countries, including the province of Quebec (e.g., Education Act of Quebec, 2000; Department for Education, 2014).

THEORETICAL FRAMEWORK

Situated Expectancy-Value Model

Decades of research have shown the usefulness of expectancy-value theory (EVT) to predict important outcomes such as career and educational aspirations (Eccles and Wigfield, 2020). According to EVT (Eccles, 1994, 2011), these outcomes stem most directly from two factors: students' expectancies of success in a given domain, and the value that they place in the domain. The expectancy component refers to the individual's self-efficacy and perceived competence, whereas the value component refers to how much they feel a task is important, worthwhile, and interesting (Wigfield and Eccles, 2000). Expectancies and values, in turn, are predicted by social and contextual influences. The most recent version of the expectancy-value model, labeled the Situated Expectancy-Value Theory (SEVT; Eccles and Wigfield, 2020), specifies that the proximal and distal aspects of the model are situation-specific and also culturally bound. In this way, the choices a student considers in a given situation are likely to be constrained by cultural values. Another feature of the SEVT is that it underlines the importance of considering both between-subjects differences and within-subjects factors to understand educational choices. Applied to choices to pursue stereotypical or counter-stereotypical career pathways, such a framework accounts for which individual factors lead students to prioritize among different domains as well as for differences between students based on factors such as gender. In the current work, we look at how parents' stereotypical beliefs shape students' individual motivation and career aspirations in the two main school domains, namely mathematics and language arts. In addition, the study compares whether these relationships differ across genders, thereby accounting for the between-person aspect of SEVT.

Parents as Transmitters of Gender Stereotypes

Parents transmit a diversity of attitudes and cultural values to their child, including gender stereotypes. In particular, different types of parental stereotypes may contribute to gender gaps in career choices and occupations. One of the most obvious forms of stereotyping relates to explicit beliefs alleging a male or female ability-superiority in domains such as mathematics and language arts (Martinot and Désert, 2007; Plante et al., 2009). In addition to such domain-specific ability stereotypes, parents may also hold stereotypical beliefs about what roles men and women should occupy in society. Specifically, such gender role beliefs may translate into conceptions that men should seek status and avoid feminine activities, or that women tend to be emotional and dependent (Sobiraj et al., 2015; Levant et al., 2017). Therefore, beyond stereotypes about ability in different domains, which are likely to affect boys' and girls' self-concepts in these domains, gender role beliefs may have implications for the types of occupational interests parents encourage or discourage, and thus make a unique contribution to students' values and aspirations toward stereotypical or counter-stereotypical domains.

Empirical work on the links between parents' gender stereotypical beliefs and students' outcomes has found that parents tend to see STEM subjects as more suitable for boys, and such beliefs are known to influence both boys' and girls' self-perceptions in mathematics and later career choices (Bleeker and Jacobs, 2004; Tomasetto et al., 2015). However, despite the fact that stereotypes associating language arts with girls are widespread in society and consistently endorsed by students (Plante et al., 2009; Chaffee et al., 2020), parental stereotypes in this domain remain understudied. One of the few studies investigating parents' gender-ability stereotypes in language arts found that, as expected, boys' expectancies and values for reading were negatively predicted by parents' stereotypes of female advantage in reading (Muntoni and Retelsdorf, 2019).

In addition, research about parents' gender role stereotypes and students' career aspirations has offered mixed results.

Specifically, although students' own gender role beliefs have been linked to motivation and aspirations, especially among boys (van der Vleuten et al., 2016; Forsman and Barth, 2017; Chaffee et al., 2020; Mastari et al., 2021), Halpern and Perry-Jenkins (2016) found no significant longitudinal association between parents' gender role beliefs and their child's gender-stereotypical occupational aspirations. In contrast, Croft et al. (2014) found that fathers' domestic gender role beliefs predicted daughters'—but not sons'—career aspirations in stereotypical domains. Such inconsistencies might be the result of differences in how gender role stereotypes relate to different school domains. For instance, McFadden et al. (2020) observed that parental gender role beliefs were more predictive of outcomes in mathematics than language arts. Although these researchers attributed their results to relatively stronger cultural mathematics than language arts ability stereotypes, such an interpretation is inconsistent with findings showing the reverse pattern (e.g., Plante et al., 2019). On the whole, these mixed results involving domain- and gender-differences highlight the need to examine how parents' stereotypes can translate into their child's motivation and career aspirations in multiple stereotyped school domains.

The Present Study

To fill this gap, the present study simultaneously considers parents' gender role beliefs and ability stereotypes. Specifically, in relying on a dyadic design including both parent and student reports to test the preregistered[1] model pictured in **Figure 1**, this study aims to develop a more complete understanding of how parents' stereotypical beliefs may influence students' motivation and decision to pursue a typical or atypical field. Another original aspect of this research is that it includes two domains that have been traditionally stereotyped as more appropriate for male (mathematics) or female (language arts) students. Such a design will help us to determine whether parents' gender stereotypes have distinct implications for boys' and girls' gendered aspirations.

We expect that parents' beliefs will relate to students' expectancies and task values for mathematics and language arts, and in turn that these motivational variables will predict students' career aspirations in these two domains. Because each domain is stereotyped in a different direction, it is expected that parents' traditional stereotypes in mathematics (i.e., stereotypes positing a male advantage), as well as their traditional gender role beliefs, will have a positive relation with their sons' mathematics expectancies and values, but a negative relation with their daughters' mathematics expectancies and values (hypothesis 1). The opposite pattern is expected for language arts, a domain that is traditionally associated with girls (hypothesis 2). In addition, because some prior work found direct links between gender role beliefs and career aspirations (Croft et al., 2014), we also expect that parent gender role stereotypes will directly relate to more gender-traditional career interests (hypothesis 3). It is also expected that the links between gender role stereotypes, motivation, and outcomes may be stronger among boys than among girls (hypothesis 4). This hypothesis is based both on previous research showing that gender norms for boys tend to be more restrictive than those for girls (Lytton and Romney, 1991; Sullivan et al., 2018), as well as on research in precarious masculinity theory. This theory suggests that masculinity is a precarious status but femininity is more stable, and that men are consequently more sensitive to gender prototypicality threats than women (Bosson and Michniewicz, 2013; Vandello and Bosson, 2013). For ability stereotypes, no specific gender differences are predicted in the strength of their relations with other variables.

[1] Preregistration at: https://osf.io/e354z/

MATERIALS AND METHODS

Participants and Procedures

This study was conducted using a subsample of 170 parent-child dyads from a larger study of students in their final year of high school. Students (60.6% girls, M_{age} = 16.15, SD = 0.45) and their parent (81.8% mothers, M_{age} = 48.03, SD = 5.81) completed questionnaires at the beginning of the school year. In four cases in which two parents completed the questionnaire for the same student, one parent was retained for analysis at random. Students were enrolled in public (35.5%) or private (64.7%) francophone schools in the metropolitan region of a large Canadian city. Almost half the students were enrolled in a regular, non-selective school track (49.4%), 11.7% were in other non-selective programs such as arts or physical education programs, and 31.8% were in enriched selective school tracks that included advanced mathematics instruction. Students from all school tracks and programs were enrolled in daily mathematics and language arts courses. Approximately two-thirds (67.1%) of students reported their ethnicity as white or European, 10.6% as multiethnic, 5.9% as Middle Eastern or North African, 4.7% as South or Southeast Asian, 4.1% South American or Latinx, 4.1% Caribbean. Other ethnicities were reported by fewer than 5 students each. A majority of both students (82.6%) and parents (70%) reported having been born in Canada.

Students completed the questionnaires in their mathematics or language arts classrooms during regular school hours. Teachers were present during the questionnaire administration, but were asked to remain at their desks so they would not see students' responses. Research assistants read the consent forms and questionnaire items aloud to students. Students were provided with flyers, paper questionnaires, and addressed stamped envelopes to take home to their parents. Parents were invited to participate on their own time, either online via LimeSurvey or using the paper questionnaires provided to students. Parent and student participants were each offered a $10 honorarium to compensate their participation, with students' honoraria being provided to their teachers to fund a reward for the class.

Materials

Parent Ability Stereotypes

Parents reported their stereotypes about gendered ability in mathematics and language arts using a short version of the scale initially developed by Leder and Forgasz (2002) and adapted

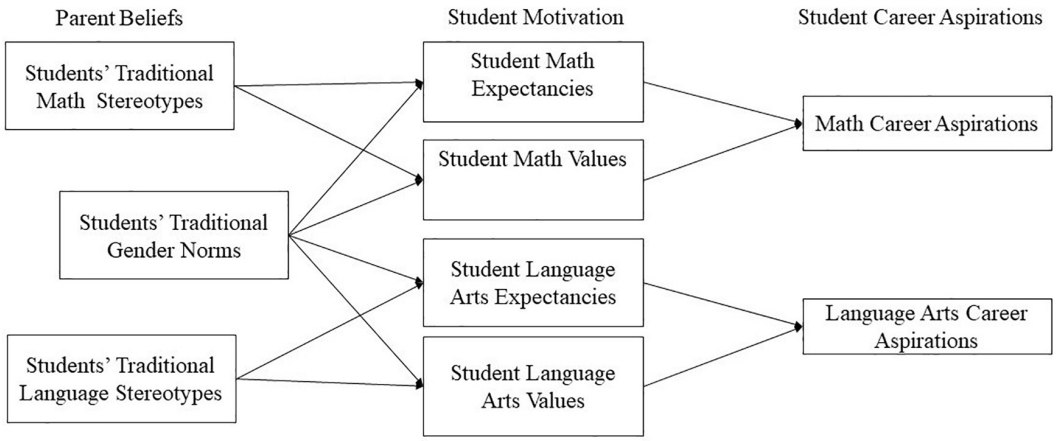


FIGURE 1 | Hypothesized model.

into French by Plante (2010). For each domain, the current measure included ten items separated into two subscales: a Male Domain scale measuring stereotypes of boys ("Boys are naturally better in mathematics/language arts"), and a Female Domain scale measuring stereotypes of girls ("Girls are naturally better in mathematics/language arts"). For each item, parents responded on a scale from 1 (strongly disagree) to 5 (strongly agree). In accordance with previous work using this measure, difference scores were calculated to reflect parents' traditional gender stereotypes in each domain. In mathematics, the subtraction [Male Domain – Female Domain] was performed for each item, whereas in language arts, the subtraction [Female Domain – Male Domain] was computed. For each domain, a higher score indicated a stronger mathematics-male or language arts-female stereotype. Internal consistency for the final scale in each domain (based on the difference scores) was high ($\omega_{mathematics}$ = 0.85; $\omega_{language\ arts}$ = 0.84).

Parent Gender Role Stereotypes

The measures of parents' gender role beliefs comprised three subscales drawn from two existing measures which were translated into French. Participants responded on a 7-point scale from 1 (do not agree at all) to 7 (agree completely). First, parents reported their beliefs about masculine gender roles using items adapted from two subscales of the Male Role Norms Scale (Thompson and Pleck, 1986). Specifically, they responded to three items reflecting masculine status-seeking ("A man owes it to his family to work at the best paying job he can get;" ω = 0.78) and three items reflecting antifemininity ("It bothers me when a man does something that I consider 'feminine'"; ω = 0.61).

Second, parents reported their beliefs about feminine gender roles using items adapted from two subscales of the Femininity Ideology Scale Short form (Levant et al., 2017). Specifically, they responded to three items reflecting emotionality ("It is expected that women will be viewed as overly emotional"; ω = 0.77) and three items reflecting dependence ("A woman should not be competitive"; ω = 0.62). These subscales were further adapted into a single composite variable based on confirmatory factor analyses (CFA) presented below. For both gender role stereotype scales, high scores indicate greater agreement with traditional gender roles.

Student Motivation

The measure of student motivation relied on two indicators for each subject: expectancies and task values. Specifically, students reported their expectancies and values in mathematics and language arts using a measure validated among Canadian students by Plante et al. (2013a; originally developed by Eccles and Wigfield, 1995). For each subject, participants responded to five items measuring expectancies of success (e.g., "How well do you think you will do in your mathematics/language arts course this year?"; $\omega_{mathematics}$ = 0.94, $\omega_{language\ arts}$ = 0.95) and six items measuring task values (e.g., "How much do you like mathematics/language arts?"; $\omega_{mathematics}$ = 0.80, $\omega_{language\ arts}$ = 0.85). For both subscales, items were rated on a 7-point scale tailored to the question wording (e.g., "very poorly" to "very well"; "not at all" to "very much"; measures can be viewed on the project's osf page[2]), with high scores indicating high levels of expectancies and values.

Student Career Aspirations

Students rated their career aspirations for jobs requiring frequent use of mathematics or language arts on a scale from 1 ("not at all true for me") to 4 ("completely true for me") using two single-item measures adapted from Crombie et al. (2005) and Stevens et al. (2007). These items were previously translated for use with French-speaking Canadian students by Plante et al. (2013a).

RESULTS

Prior to addressing our main research questions, we report the results of analysis of missing data and invariance analyses conducted to examine the psychometric equivalence of the scales across boys and girls. Then, descriptive statistics and analyses

[2]Project on osf: https://osf.io/xqr35/

TABLE 1 | Measurement invariance by gender.

	χ^2	df	*p*	RMSEA	CFI	TLI	Δdf	Δ SB χ^2
Masculinity beliefs								
Configural	45.21	16	0.000	0.150	0.816	0.654		
Metric	42.57	22	0.005	0.107	0.870	0.823	6	3.99
Scalar	45.23	26	0.011	0.095	0.878	0.860	4	2.31
Strict	42.64	32	0.099	0.064	0.933	0.937	6	3.00
Covariances	41.69	33	0.143	0.057	0.945	0.950	1	0.05
Femininity beliefs								
Configural	7.25	4	0.123	0.100	0.962	0.886		
Metric	11.97	8	0.153	0.078	0.954	0.931	4	6.96
Scalar	14.63	11	0.200	0.064	0.958	0.954	3	1.51
Strict	23.61	15	0.072	0.084	0.900	0.920	4	7.38
Traditional ability stereotypes								
Configural	105.63	66	0.001	0.086	0.903	0.867		
Metric	109.80	76	0.007	0.074	0.917	0.902	10	6.83
Scalar	119.74	84	0.006	0.072	0.912	0.906	8	9.21
Strict	136.13	94	0.003	0.074	0.897	0.901	10	15.80
Covariances	133.60	96	0.007	0.069	0.908	0.914	2	0.55
Expectancy-values in language arts								
Configural	124.10	80	0.001	0.081	0.956	0.939		
Metric	133.10	91	0.003	0.074	0.958	0.949	11	9.26
Scalar	147.60	102	0.002	0.073	0.954	0.950	11	14.72
Strict	226.17	113	0.000	0.109	0.886	0.889	11	80.37***
Partial strict	151,29	107	0.003	0.070	0.955	0.954	5	3.33
Covariances	170.83	111	0.000	0.080	0.940	0.940	4	16,85**
Partial covariance	154.92	110	0.003	0.070	0.955	0.955	3	3.70
Expectancy-values in math								
Configural	154.73	82	0.000	0.102	0.931	0.908		
Metric	181.90	93	0.000	0.106	0.916	0.900	11	27.28**
Partial metric	171.62	92	0.000	0.101	0.925	0.910	10	16.87
Scalar	211.55	101	0.000	0.114	0.895	0.886	9	49.16***
Partial scalar	185.04	99	0.000	0.101	0.919	0.910	7	13.48
Strict	190.42	110	0.000	0.093	0.924	0.924	11	6.79
Covariances	191.86	113	0.000	0.091	0.925	0.927	3	1.49

*Satorra–Bentler (SB) scaling is used for χ^2 difference tests comparing nested models. **$p < 0.01$, ***$p < 0.001$.*

of mean gender differences are presented. Finally, we report the results of the hypothesized model, tested using path analysis with latent factor scores, and of model comparisons by gender.

Preliminary Analyses

Examination of the data revealed that missing data ranged from 4 to 5%. In addition, the non-significant result of Little's test [$\chi^2(33) = 26.92$, $p = 0.763$] suggested that missingness was completely at random. Therefore, full information maximum likelihood was used to address missing data in MPlus (Muthén and Muthén, (1998-2017)) using the MLR estimator.

The measurement invariance of each scale was evaluated in a series of CFAs using nested models to test the equivalence of configural, metric, scalar, strict, and (where applicable) covariances and correlated uniqueness models across boys and girls (see **Table 1**). Measurement experts suggest that comparisons of latent means are supported for variables showing at least full scalar invariance (Putnick and Bornstein, 2016). Factors showing at least partial invariance (with fewer than half of the parameters non-invariant) are also commonly accepted, as simulation studies suggest partial invariance is likely to result in minimal bias (Hsiao and Lai, 2018). For femininity beliefs, because a two-factor solution including separate latent factors for emotionality and dependence fit poorly [$\chi^2(16) = 71.61$, $p < 0.001$, CFI = 0.78, RMSEA = 0.21], a single factor was computed. Two problematic emotionality items were removed, resulting in one latent femininity ideology factor with four indicators. Following this modification, all variables showed acceptable levels of measurement invariance, supporting comparisons of means and models by gender. Expectancies and values in language arts showed only partial strict invariance, with two factor loadings freed for expectancies and two for values, and partial invariance of correlated uniquenesses, with one inter-item correlation freed. In mathematics, task values showed only partial metric invariance, with one factor loading freed, and partial scalar invariance with two item intercepts

TABLE 2 | Means and standard deviations of observed variables by student gender.

		Girls		Boys		Overall	
	Range	Mean	*SD*	Mean	*SD*	Total	*SD*
Parent-reported variables							
Math stereotype	−5 – 5	0.33	0.71	0.64	0.87	0.45	0.79
Language stereotype	−5 – 5	0.70	0.73	0.84	0.81	0.75	0.76
Masculine status-seeking	1 – 7	2.69	1.45	3.18	1.61	2.89	1.53
Masculine antifemininity	1 – 7	1.72	0.92	1.88	1.18	1.77	1.03
Feminine emotionality and dependence	1 – 7	1.47	0.71	1.43	0.63	1.46	0.68
Student-reported variables							
Math expectancies	1 – 7	4.99	1.25	5.19	1.24	5.06	1.25
Math values	1 – 7	4.90	1.02	4.90	1.25	4.90	1.11
Language arts expectancies	1 – 7	5.30	1.07	4.26	1.14	4.90	1.20
Language arts values	1 – 7	5.61	0.88	4.41	1.16	5.15	1.15
Math career aspirations	1 – 4	2.37	0.98	2.86	1.04	2.56	1.03
Language career aspirations	1 – 4	2.50	0.99	1.80	0.95	2.23	1.03

TABLE 3 | Correlations by gender.

	1	2	3	4	5	6	7	8	9	10	11
(1) Math stereotype		0.64**	−0.10	−0.02	0.0	0.00	0.06	−0.15	−0.09	0.15	−0.18
(2) Language stereotype	0.34		−0.33**	−0.13	−0.07	−0.17	−0.12	0.05	0.14	0.16	0.06
(3) Masc. status	−0.01	0.09		0.74***	0.28*	0.04	0.15*	−0.12	−0.32**	−0.03	−0.24*
(4) Masc. antifemininity	0.01	0.10	0.53***		0.40**	0.01	0.17*	−0.02	−0.19	0.03	−0.08
(5) Fem. emotionality and dependence	0.01	0.06	0.39***	0.29***		−0.02	0.08	−0.08	−0.19	0.06	−0.07
(6) Math expectancies	0.15**	0.06	0.09	0.05	0.06		0.65***	−0.05	−0.11	0.29*	−0.23
(7) Math values	−0.02	−0.04	0.13	0.09	0.07	0.56***		−0.02	−0.12	0.62***	−0.30
(8) Language expectancies	0.17*	0.22**	0.03	0.04	−0.10	0.25*	0.20*		0.70***	−0.07	0.47***
(9) Language values	0.12	0.20**	0.09	0.11	−0.00	0.11	0.14	0.57***		−0.11	0.51***
(10) Math career asp.	−0.09	−0.07	0.08	0.05	0.04	0.17**	0.28***	−0.09	−0.16*		−0.19
(11) Language career asp.	0.05	0.14*	−0.05	0.00	−0.06	−0.04	−0.15	0.20*	0.40***	−0.31**	

*Results for girls are shown below the diagonal, and results for boys are shown above the diagonal. *p < 0.05, **p < 0.01, ***p < 0.001.*

freed. Therefore, as recommended in cases of partial invariance (Putnick and Bornstein, 2016), mean gender comparisons in value for mathematics are conducted at the latent level yet should be interpreted with caution.

Descriptive Statistics and Mean Difference Analyses

After examining mean descriptive statistics for the observed variables, reported in **Table 2**, further analyses were conducted to determine the direction of parents' stereotypes as well as to test for mean gender differences.

To determine whether parents held explicit stereotypes about mathematics and language arts in the expected directions, one-sample *t*-tests were conducted in SPSS to examine whether their stereotypes differed from the neutral midpoint of 0. The results showed that parents held stereotypes advantaging male students in mathematics [$t(163) = 7.30$, $p < 0.001$, $d = 0.57$] and female students in language arts [$t(163) = 12.62$, $p < 0.001$, $d = 0.99$], with the language arts stereotypes having the larger effect size.

Intercorrelations among the latent variables were examined (**Table 3**), and a set of analyses examined whether parents' beliefs and students' expectancies, values, and aspirations varied by student gender. Invariance testing to examine differences between latent means was conducted in MPlus. The results showed that parental stereotypes did not differ between parents of boys and parents of girls [ability stereotypes, $\Delta SB\ \chi^2(2) = 3.96$, $p = 0.138$; masculinity beliefs $\Delta SB\ \chi^2(2) = 4.14$, $p = 0.127$; femininity beliefs, $\Delta SB\ \chi^2(1) = 0.10$, $p = 0.756$]. Furthermore, girls reported significantly higher expectancies and values in language arts than boys [$\Delta SB\ \chi^2(2) = 48.20$, $p < 0.001$], but expectancies and values in mathematics showed no mean gender differences [$\Delta SB\ \chi^2(2) = 2.17$, $p = 0.338$]. Independent samples *t*-tests comparing students' career aspirations showed that boys reported higher career aspirations in mathematics than girls [$t(166) = -3.09$, $p = 0.002$, $d = 0.49$], whereas girls reported higher language arts career aspirations than boys [$t(167) = 4.61$, $p < 0.001$, $d = 0.71$].

Direct and Indirect Relations Between the Studied Variables

To further examine the relations between parents' beliefs and students' motivation and career aspirations in mathematics and language arts, latent factor scores were extracted from the most invariant measurement models. These scores were then used to

compute path models accounting for the nested nature of the data using the TYPE = COMPLEX command in MPlus. MPlus code for the models is included in the online **Supplementary Materials**, along with covariance matrices for reproducibility. The initial model fit poorly [$\chi^2(14) = 262.52$, $p < 0.001$, CFI = 0.34, RMSEA = 0.33, SRMR = 0.17]. Based on the modification indices, intercorrelations between expectancies and values and cross-domain regression paths between stereotypes, expectancies, values, and aspirations were added to the model, as these links were theoretically grounded (Eccles and Wigfield, 2020; Plante et al., 2013b). Model comparisons showed that this model was non-invariant across student gender [$\Delta\chi^2(3) = 8.86$, $p = 0.031$], suggesting that the pattern of results differed for boys and girls. Consequently, models were examined separately by gender. The final multigroup model showed a good fit to the data [$\chi^2(6) = 11.68$, $p = 0.070$, CFI = 0.99, RMSEA = 0.11, SRMR = 0.05] based on most indices. Although it should be noted that the RMSEA was above the recommended value (Browne and Cudeck, 1993), simulation studies suggest that the RMSEA often inappropriately indicates poor fit in models with low degrees of freedom (Kenny et al., 2015). Therefore, given that the chi-square, an exact fit test, was non-significant, and other indicators also suggested good fit, we retained this as our final model. The final models are pictured in **Figures 2**, **3**, respectively, for girls and boys. Because bootstrapping cannot be combined with TYPE = COMPLEX, confidence intervals for the indirect effects were computed using the Monte Carlo method with 1,000 repetitions, using the method recommended by Selig and Preacher (2008).

As can be seen in **Figure 2**, parents' stereotypes that girls are advantaged in language arts predicted stronger expectancies and values for girls in this domain, supporting hypothesis 2. In turn, girls' language arts values predicted career interest positively in language arts and negatively in mathematics. Furthermore, in accordance with the mediational prediction in hypothesis 2, girls' language arts values mediated the relation between parents' traditional language stereotypes and girls' career aspirations in language arts ($\beta_{indirect} = 0.18$, 95% CI [0.045, 0.380]). Results also revealed an un-hypothesized mediation from language arts stereotypes to mathematics career aspirations through language arts values ($\beta_{indirect} = -0.08$, 95% CI [−0.197, −0.010]). Unexpectedly, girls' language arts expectancies negatively predicted language arts career interest (indirect effect of language arts stereotypes via expectancies: $\beta = -0.06$, 95% CI [−0.188, −0.002]). Although surprising, this result apparently reflects a suppressor effect. Such effects occur when the direction of a correlation between two variables changes after controlling for other variables (e.g., Lutz, 1983). In the current case, despite a positive bivariate correlation between language arts expectancies and career aspirations ($r = 0.20$, $p = 0.029$, see **Table 3**), this link became negative in our final model. Additionally, in mathematics, girls' values predicted not only higher career aspirations in mathematics, but also lower career aspirations in language arts, partially supporting hypothesis 2. What is more surprising is that, counter to hypothesis 1, parents' mathematics ability stereotypes did not predict girls' task values in this domain. Our results also showed that girls' expectancies in mathematics positively predicted their intention to pursue a career in language arts. Again, this finding could reflect a suppressor effect, as the bivariate correlation between these variables was non-significant, as can be seen in **Table 3**. Furthermore, also contrary to hypothesis 1, the more parents reported traditional stereotypes in mathematics, the stronger girls' expectancies of success in this domain. Given that the bivariate correlation between these variables was also positive, this result cannot be attributed to a suppressor effect. In addition, the indirect effect of mathematics stereotypes on language arts interest through math expectancies was not statistically significant ($\beta_{indirect} = 0.04$, 95% CI [−0.001, 0.108]). Finally, contrary to hypothesis 3, parents' beliefs about traditional gender roles were unrelated to girls' motivation and aspirations.

For boys (**Figure 3**), the results presented a quite different pattern. Overall, hypotheses 1 and 2 were not supported among boys. Specifically, most stereotypical parent beliefs predicted boys' motivation and career aspirations in at least one of the two domains. However, none of the relations between parent beliefs and career aspirations was mediated through expectancies and task values in either language arts or mathematics. Parents' mathematics ability stereotypes were associated with lower expectancies and values in language arts among boys but were unrelated to career aspirations. Furthermore, parental beliefs that language is for girls—but not the opposite belief that mathematics is for boys—directly predicted stronger aspirations for mathematics-related careers among boys. Although these results do not support hypotheses 1 and 2, they offer an interesting alternative explanation that mathematics stereotypes may be important for understanding boys' motivation and underrepresentation in language arts. Parents' stereotypes disadvantaging boys in language arts were also associated with weaker mathematics expectancies for boys.

In addition, parents' beliefs about traditional feminine gender roles were associated with lower expectancies and values for language arts among boys but not girls, providing support for hypothesis 4 and partial support for hypothesis 3. Specifically, parents' beliefs that men should seek status were related to lower value for language arts, and also to less interest in careers using language skills, which is consistent with hypothesis 3. Less expectedly, and contrary to hypothesis 3, parents' beliefs that men should avoid activities that appear feminine were related to greater language career aspirations and language arts value among boys. Once again, these counterintuitive results appear to reflect a suppressor effect since the bivariate correlations between these variables were negative (see **Table 3**). It is also noteworthy that although boys' interest in mathematics careers was predicted by their value for mathematics, no other expectancy-value variables predicted their career aspirations. Instead, boys' interest in language arts careers was predicted directly and exclusively by their parents' gender role stereotypes. This last result partially supports hypothesis 3 and brings interesting insights about gender-specificity in the mechanisms by which parental stereotypes may influence boys' and girls' language arts interests differently.

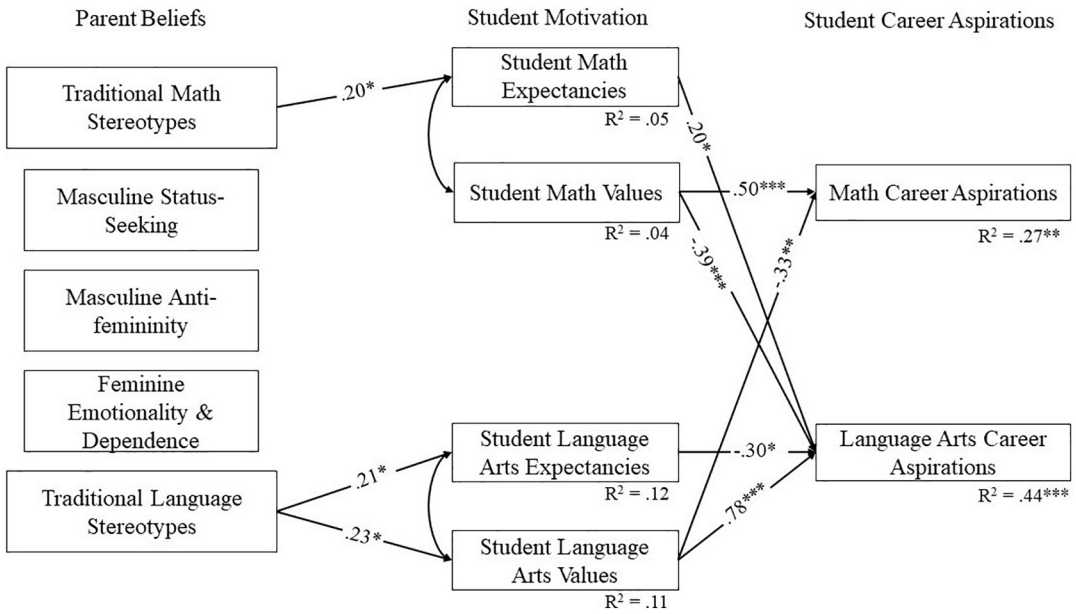


FIGURE 2 | Final model for girls showing statistically significant paths. Standardized coefficients are shown. All error terms are significant at $p < 0.001$. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

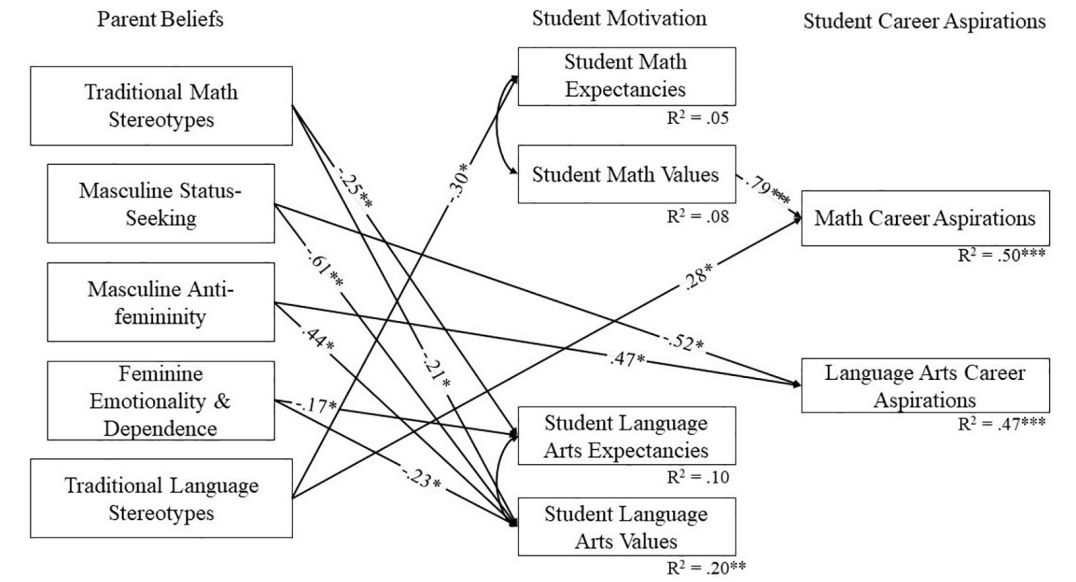


FIGURE 3 | Model for boys showing statistically significant paths. Standardized coefficients are shown. All error terms are significant at $p < 0.001$. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

DISCUSSION

Our research showed that parents, as socializers who hold a variety of gender stereotypical beliefs, may have a key role especially at the end of high school, a critical period during which students must choose between multiple domains as they enter either higher education or the workforce. Specifically, in using a dyadic design, this research provided original insights about possible mechanisms by which parents might influence their child's career aspirations toward stereotypical or counter-stereotypical domains such as mathematics or language arts. Furthermore, the study extended prior findings in showing interesting gender and school domain differences in the processes by which parental beliefs relate to students' motivation and career interests. These results have both theoretical and practical implications.

Understanding Career Aspirations for Boys and Girls

In studying parents' stereotypical beliefs, our data shed light on the potential socialization processes through which gender imbalances emerge. Our results showed that the ways by which parents' beliefs relate to students' career aspirations are quite different for boys and girls. For girls, our results supported hypothesis 2 that parents' ability stereotypes advantaging girls in language arts were related to their daughters' career aspirations through their motivational beliefs in this domain. For boys, results instead showed that when parental beliefs were associated with students' career aspirations, the link was direct. This finding is consistent with past work showing that, especially among boys, student or peer gender role beliefs are directly linked to occupational interests (van der Vleuten et al., 2016; Mastari et al., 2021). For girls, however, these links have been found to be fully mediated by motivational beliefs, as expected under SEVT (Plante et al., 2013a). This is interesting in light of the fact that the expectancy-value model was initially developed and tested mainly in the context of understanding female students' underrepresentation in STEM fields (Eccles, 1994; Eccles and Wigfield, 2020). Although SEVT is expected to apply to students of any gender, our results suggest that relations between stereotypes and career aspirations may not always be mediated through expectancies or values among boys. Therefore, to increase our theoretical understanding and to guide interventions, future research focusing on gender differences is needed to better capture the processes through which gender stereotypes influence boys' and girls' career decision-making.

Another interesting finding highlighted by the current study is that multiple types of parental gender beliefs related directly to either boys' career aspirations or motivation, partially supporting hypothesis 3. In particular, parents' gender role beliefs, or their beliefs about how men and women should behave, were influential exclusively for boys, supporting hypothesis 4. This finding is consistent with research suggesting that gender role norms tend to be more restrictive for boys than for girls (Sullivan et al., 2018), but it additionally highlights that prescriptive gender role norms might contribute to adolescent boys' educational and occupational decision-making. Interestingly, parental belief in feminine gender roles was related to lower expectancies and values in language arts among boys; surprisingly, however, such beliefs did not lead to lower language arts career aspirations. In fact, boys' language arts career aspirations were predicted only by their parents' beliefs about masculine gender roles such that boys with parents who more strongly believed that it is important for men to seek high status were particularly uninterested in language-related careers. What is less intuitive is the finding that after controlling for these status beliefs, parents' beliefs that men should avoid femininity related to stronger language arts career aspirations for boys. Although the particular processes explaining this suppressor effect are unclear, this result suggests that these two facets of masculine gender role stereotypes (i.e., status-seeking and antifemininity) did not additively contribute to predicting boys' aspirations toward language fields. Nonetheless, such results are particularly informative as they go beyond previous work using more general measures of gender normative stereotypes (Croft et al., 2014; McFadden et al., 2020) and indicate that different facets of gender role beliefs might have distinct implications for boys' motivation and career aspirations in different domains.

Despite the importance of gender role stereotypes for boys, parents' traditional ability stereotypes did not predict boys' career aspirations in language arts. However, boys whose parents reported traditional language arts stereotypes reported more interest in mathematical careers and, surprisingly, lower expectancies of success in mathematics. Though it is not surprising that parents' negative stereotypes about boys might negatively relate to their sons' motivation, it is surprising that this result was observed in mathematics rather than in language arts. Interestingly, the hypothesis that traditional mathematics stereotypes would boost boys' mathematics motivation and career aspirations (hypothesis 1) was not supported by our results. Instead, parents' stereotypes advantaging boys in this domain were associated with boys' devaluation of language arts, as well as with lower expectancies of success in language arts. Together, these findings could be explained by the fact that even though parents still hold mathematics stereotypes advantaging boys, students themselves do not, as shown by a growing body of research on explicit stereotypes (e.g., Schmader et al., 2004; Martinot and Désert, 2007; Kurtz-Costes et al., 2014). Furthermore, prior research has shown that students' neutral or even female-advantaging stereotypes in mathematics were internalized through students' expectancies and task values in mathematics (e.g., Plante et al., 2013a). In other words, boys' own stereotypes might mitigate the role of their parents' beliefs in mathematics, whereas parental stereotypes may still contribute to the devaluation of competing domains such as language arts.

For girls, contradicting hypothesis 1, parents' mathematics ability stereotypes did not relate to lower motivational indicators in mathematics, nor to lower mathematics career aspirations. Rather, as a result of a suppressor effect, girls' higher expectancies in mathematics were related to stronger language arts career aspirations. In addition, parents' mathematics ability stereotypes were positively related to mathematics expectancies, a relation that was also observed in the bivariate correlations. One possible explanation for this result is that parenting a mathematically gifted daughter might make stereotypes about girls and mathematics more salient, leading parents of such daughters to report stronger stereotypes in this domain. Another possibility is that parents who hold traditional beliefs in mathematics may devote additional support to help their daughters succeed in mathematics in the hope of counteracting these stereotypes.

Domain Specificities in the Development of Career Aspirations

The current study underlined different patterns both in mean differences and in the relations between parental beliefs and student variables across the domains of mathematics and language arts. First, in terms of mean differences, this study showed that parents held traditional stereotypes in both domains. Gender differences in students' motivational beliefs

were consistent with their parents' stereotypical conceptions in language arts but not in mathematics, as gender differences in expectancies and values were observed only in language arts. Such findings could be explained by the fact that interventions to reduce stereotypes of mathematics may have been effective in reducing gender gaps between boys' and girls' motivation in mathematics, but without reaching parents, who still hold more old-fashioned stereotypes. This interpretation is aligned with work showing that explicit mathematics stereotypes are fading among students, while language arts stereotypes remain consistent (Plante et al., 2009, 2019). In contrast, parents' conceptions in our sample were surprisingly similar to those in a seminal study conducted 30 years ago showing that parents endorsed traditional stereotypes in mathematics (Jacobs, 1991).

Second, in terms of relations between parental stereotypes and student variables, our study showed that it was primarily language arts stereotypes that were predictive of adolescents' career aspirations. For girls, the more parents stereotyped language arts as female-advantaged, the more girls were motivated in language arts and interested in language arts careers. For boys, disadvantaging language arts stereotypes were directly related to stronger mathematics career aspirations. In other words, these results could mean that parents' language arts stereotypes did not discourage their son's interest in language arts careers, but rather attracted them to mathematical careers, a hypothesis that needs to be empirically supported. On the other hand, hypothesis 1, that parents' mathematics ability stereotypes would relate to students' motivation and career aspirations, was unsupported. Instead, parents' ability stereotypes in mathematics may have undermined boys' motivation toward language arts in school. Based on these results, language arts stereotypes may be more influential than mathematics stereotypes in predicting students' career interests and therefore should receive greater attention.

The current study also showed interesting cross-domain processes that could help researchers understand career aspirations and career choices. Consistent with previous work on the topic (Wang, 2012; Plante et al., 2013a), girls' task values in both mathematics and language arts were strongly related to career aspirations in the corresponding domain. Less expectedly, task values were also negatively related to girls' career aspirations in the competing domain. Such findings further support the importance of considering students' relative valuation of different domains (Chow and Salmela-Aro, 2011; Plante et al., 2019; Eccles and Wigfield, 2020). For instance, even if girls highly value mathematics, a higher valuation of language arts could not only still lead them to a language arts career path, but also decrease their aspirations toward a mathematical career. For boys, however, such cross-domain results involving task values were not observed. Rather, aside from parents' traditional language arts stereotypes, only boys' task values in mathematics predicted their aspirations in that domain. Unexpectedly, none of the motivational beliefs in language arts related to boys' career aspirations. Based on these results, it appears that girls consider both mathematics and language arts careers as valuable options, and that their motivational beliefs toward these two competing domains might have a complementary role in shaping their aspirations. For boys, it instead seems that parents' stereotypical beliefs could contribute to push them away from counter-stereotypical careers such as language arts fields, leaving mathematics as their only valued option.

Limitations and Future Directions

The present study has some limitations that should be acknowledged when interpreting the results. First, this study's use of path analysis based on correlational data and relying on a single measurement timepoint for students' indicators prevents us from drawing causal inferences. A second limitation is that our sample consisted mainly of mother/child dyads. Therefore, the role of fathers' gender stereotypes in students' motivation and career aspirations may be attenuated in our results. Our use of path analysis also revealed a few suppressor effects that were difficult to explain. Despite these limitations, the dyadic nature of the data provides convincing evidence that parents' beliefs relate to their son's or daughter's motivation and career interests. Nonetheless, such findings need to be further replicated using experimental designs to clearly establish causal links among these variables. For instance, the effects of interventions to change parents' stereotypical conceptions on students' career aspirations and actual course enrollment decisions would be a valuable avenue for future research.

In addition, although the current study was innovative in modeling two school domains together, its generalizability is limited to these two domains. Thus, it is difficult to determine whether boys believe all non-STEM domains are uninteresting, or if such beliefs only apply to language arts. Furthermore, because real-world career decision-making involves choosing among more than two domains, future research simultaneously including a wider variety of domains would be useful to better assess the ecological validity of the results.

CONCLUSION

In examining parent-student dyads, our research suggested that parents, as important socializers, could transmit stereotypes that predict students' motivation and career aspirations. Furthermore, by simultaneously measuring different types of gender stereotypical beliefs among parents and considering two school domains, our study showed that these processes are both gender and domain specific. In summary, for girls, our findings suggest that parents' ability stereotypes about language might boost girls' motivation for language arts, thereby nudging them away from STEM pathways. Our results also provide further evidence that girls' career choices stem not only from their valuation of the corresponding domain, but also from their valuation of competing domains. Such findings highlight the need to consider multiple domains simultaneously to better capture the complexity of girls' career decisions. Meanwhile, for boys, parents' language ability stereotypes were directly related to mathematical career aspirations, and their mathematics ability stereotypes related to poorer motivation in language arts among boys. These results suggest that stereotypes that mathematics is for boys and language arts is for girls might push boys away

from language arts and toward mathematics. Our study also highlighted the unique role of parental beliefs in traditional gender roles for boys' motivation and career aspirations. Specifically, parents' gender role stereotypes directly related to less interest in language arts only among boys, thus pointing to an important avenue for future research into gender gaps in female-dominated fields. Taken together, these domain- and gender-specific results could guide interventions to promote gender equity not only in traditionally male-dominated, mathematics-heavy fields, but also in female-dominated language fields.

DATA AVAILABILITY STATEMENT

The anonymized data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comité Institutionnel D'éthique De La Recherche Avec Des Êtres Humains at the Université du Québec à Montréal. Participants provided written informed consent for their own participation; written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with provincial legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

KC and IP contributed to conception and design of the study and wrote the manuscript. KC performed the statistical analysis. Both authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

Preparation of this manuscript was supported by funding from the Fonds de recherché du Québec, specifically MEES/FRQSC grant number 264571 and FRQSC postdoctoral award number 283864, as well as by funding from the Social Sciences and Humanities Research Council of Canada, grant number SSHRC 430-2018-00757.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.796073/full#supplementary-material

REFERENCES

Bandura, A. (1977). Self-efficacy: towards a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215. doi: 10.1037/0033-295X.84.2.191

Bleeker, M. M., and Jacobs, J. E. (2004). Achievement in Math and Science: do Mothers' Beliefs Matter 12 Years Later? *J. Educ. Psychol.* 96, 97–109. doi: 10.1037/0022-0663.96.1.97

Bosson, J. K., and Michniewicz, K. S. (2013). Gender dichotomization at the level of ingroup identity: what it is, and why men use it more than women. *J. Pers. Soc. Psychol.* 105, 425–442. doi: 10.1037/a0033126

Browne, M. W., and Cudeck, R. (1993). "Alternative ways of assessing model fit," in *Testing Structural Equation Models*, eds K. A. Bollen and J. S. Long. Newbury Park: Sage), 136–162.

Chaffee, K. E., Lou, N. M., Noels, K. A., and Katz, J. W. (2020). Why don't "real men" learn languages? Masculinity threat and gender ideology suppress men's language learning motivation. *Group Process. Intergroup Relat.* 23, 301–318. doi: 10.1177/1368430219835025

Chow, A., and Salmela-Aro, K. (2011). Task-values across subject domains: a gender comparison using a person-centered approach. *Int. J. Behav. Dev.* 35, 202–209. doi: 10.1177/0165025411398184

Croft, A., Schmader, T., and Block, K. (2015). An underexamined inequality: cultural and psychological barriers to men's engagement with communal roles. *Pers. Soc. Psychol. Rev.* 19, 343–370. doi: 10.1177/1088868314564789

Croft, A., Schmader, T., Block, K., and Baron, A. S. (2014). The second shift reflected in the second generation: do parents' gender roles at home predict children's aspirations? *Psychol. Sci.* 25, 1418–1428. doi: 10.1177/0956797614533968

Crombie, G., Sinclair, N., Silverthorn, N., Byrne, B. M., DuBois, D. L., and Trinneer, A. (2005). Predictors of Young Adolescents? Math Grades and Course Enrollment Intentions: gender Similarities and Differences. *Sex Roles* 52, 351–367. doi: 10.1007/s11199-005-2678-1

Department for Education (2014). *The National Curriculum in England: Complete Framework for Key Stages 3 and 4.* Available online at: https://www.gov.uk/government/publications/national-curriculum-in-england-secondary-curriculum

Eccles, J. S. (1994). Understanding women's educational and occupational choices. *J. Women Q.* 18, 585–609. doi: 10.1111/j.1540-4560.2011.01718.x

Eccles, J. S. (2011). Gendered educational and occupational choices: applying the Eccles et al. Model of achievement-related choices. *Int. J. Behav. Dev.* 35, 195–201. doi: 10.1177/0165025411398185

Eccles, J. S., and Wigfield, A. (1995). In the mind of and actor: the structure of adolescents' academic achievement related-beliefs and self-perceptions. *Pers. Soc. Psychol. Bull.* 21, 215–225. doi: 10.1177/0146167295213003

Eccles, J. S., and Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. *Contemp. Educ. Psychol.* 61:101859. doi: 10.1016/j.cedpsych.2020.101859

Education Act of Quebec (2000). *Chapter I-13.3, r. 8 Basic School Regulation for Preschool, Elementary and Secondary Education.* Available online at: http://www.legisquebec.gouv.qc.ca/fr/document/rc/I-13.3,%20r.%208

Forsman, J. A., and Barth, J. M. (2017). The Effect of Occupational Gender Stereotypes on Men's Interest in Female-Dominated Occupations. *Sex Roles* 76, 460–472. doi: 10.1007/s11199-016-0673-3

Froehlich, L., Olsson, M. I. T., Dorrough, A. R., and Martiny, S. E. (2020). Gender at Work Across Nations: men and Women Working in Male-Dominated and Female-Dominated Occupations are Differentially Associated with Agency and Communion. *J. Soc. Issues* 76, 484–511. doi: 10.1111/josi.12390

Halpern, H. P., and Perry-Jenkins, M. (2016). Parents' Gender Ideology and Gendered Behavior as Predictors of Children's Gender-Role Attitudes: a Longitudinal Exploration. *Sex Roles* 74, 527–542. doi: 10.1007/s11199-015-0539-0

Hsiao, Y.-Y., and Lai, M. H. C. (2018). The Impact of Partial Measurement Invariance on Testing Moderation for Single and Multi-Level Data. *Front. Psychol.* 9:740. doi: 10.3389/fpsyg.2018.00740

Hyde, J. S. (2014). Gender similarities and differences. *Annu. Rev. Psychol.* 65, 373–398. doi: 10.1146/annurev-psych-010213-115057

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., and Hopp, C. (1990). Gender Comparisons of Mathematics Attitudes and Affect: a Meta-Analysis. *Psychol. Women Q.* 14, 299–324. doi: 10.1111/j.1471-6402.1990.tb00022.x

Jacobs, J. E. (1991). Influence of gender stereotypes on parent and child mathematics attitudes. *J. Educ. Psychol.* 83, 518–527. doi: 10.1037/0022-0663.83.4.518

Kenny, D. A., Kaniskan, B., and McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Soc. Methods Res.* 44, 486–507. doi: 10.1177/0049124114543236

Kurtz-Costes, B., Copping, K. E., Rowley, S. J., and Kinlaw, C. R. (2014). Gender and age differences in awareness and endorsement of gender stereotypes about academic abilities. *Eur. J. Psychol. Educ.* 29, 603–618. doi: 10.1007/s10212-014-0216-7

Leder, G. C., and Forgasz, H. J. (2002). *Two New Instruments to Probe Attitudes About Gender and Mathematics*. Melbourne Vic Australia: La Trobe University, 28.

Levant, R. F., Alto, K. M., McKelvey, D. K., Richmond, K. A., and McDermott, R. C. (2017). Variance composition, measurement invariance by gender, and construct validity of the femininity ideology scale-short form. *J. Couns. Psychol.* 64, 708–723. doi: 10.1037/cou0000230

Lutz, J. G. (1983). A Method for Constructing Data Which Illustrate Three Types of Suppressor Variables. *Educ. Psychol. Meas.* 43, 373–377. doi: 10.1177/001316448304300206

Lytton, H., and Romney, D. M. (1991). Parents' differential socialization of boys and girls: a meta-analysis. *Psychol. Bull.* 109, 267–296. doi: 10.1037/0033-2909.109.2.267

Martinot, D., and Désert, M. (2007). Awareness of a gender stereotype, personal beliefs and self-perceptions regarding math ability: when boys do not surpass girls. *Soc. Psychol. Educ.* 10, 455–471. doi: 10.1007/s11218-007-9028-9

Mastari, L., Spruyt, B., and Siongers, J. (2021). Preadolescents' Stereotypical Occupational Preferences: a Matter of Competence or Culture? *Child Indic. Res.* 14, 1199–1225. doi: 10.1007/s12187-021-09802-7

McFadden, K. E., Puzio, A., Way, N., and Hughes, D. (2020). Mothers' gender beliefs matter for adolescents' academic achievement and engagement: an examination of ethnically diverse U.S. mothers and adolescents. *Sex Roles* 84, 166–182. doi: 10.1007/s11199-020-01157-7

Muntoni, F., and Retelsdorf, J. (2019). At their children's expense: how parents' gender stereotypes affect their children's reading outcomes. *Learn. Instr.* 60, 95–103. doi: 10.1016/j.learninstruc.2018.12.002

Muthén, L. K., and Muthén, B. O. (1998-2017). *Mplus User's Guide (Eighth Edition)*. Los Angeles: Muthén & Muthén.

Olsson, M., and Martiny, S. E. (2018). Does Exposure to Counterstereotypical Role Models Influence Girls' and Women's Gender Stereotypes and Career Choices? A Review of Social Psychological Research. *Front. Psychol.* 9:2264. doi: 10.3389/fpsyg.2018.02264

Plante, I. (2010). Adaptation et validation d'instruments de mesure des stéréotypes de genre en mathématiques et en français. *Mes. Éval. Éduc.* 33, 1–34. doi: 10.7202/1024894ar

Plante, I., De la Sablonnière, R., Aronson, J. M., and Théorêt, M. (2013a). Gender stereotype endorsement and achievement-related outcomes: the role of competence beliefs and task values. *Contemp. Educ. Psychol.* 38, 225–235. doi: 10.1016/j.cedpsych.2013.03.004

Plante, I., O'Keefe, P. A., and Théorêt, M. (2013b). The relation between achievement goal and expectancy-value theories in predicting achievement-related outcomes: a test of four theoretical conceptions. *Motiv. Emot.* 37, 65–78. doi: 10.1007/s11031-012-9282-9

Plante, I., O'Keefe, P. A., Aronson, J., Fréchette-Simard, C., and Goulet, M. (2019). The interest gap: how gender stereotype endorsement about abilities predicts differences in academic interests. *Soc. Psychol. Educ.* 22, 227–245. doi: 10.1007/s11218-018-9472-8

Plante, I., Théorêt, M., and Favreau, O. E. (2009). Student gender stereotypes: contrasting the perceived maleness and femaleness of mathematics and language. *Educ. Psychol.* 29, 385–405. doi: 10.1080/01443410902971500

Putnick, D. L., and Bornstein, M. H. (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev. Rev.* 41, 71–90. doi: 10.1016/j.dr.2016.06.004

Schmader, T., Johns, M., and Barquissau, M. (2004). The costs of accepting gender differences: the role of stereotype endorsement in women's experience in the math domain. *Sex Roles* 50, 835–850. doi: 10.1023/b:sers.0000029101.74557.a0

Selig, J. P., and Preacher, K. J. (2008). *Monte Carlo Method for Assessing Mediation: An Interactive Tool for Creating Confidence Intervals for Indirect Effects [Computer Software]*. Available online at: http://quantpsy.org/

Simon, R. M., Wagner, A., and Killion, B. (2016). Gender and choosing a STEM major in college: femininity, masculinity, chilly climate, and occupational values. *J. Res. Sci. Teach.* 54, 299–323. doi: 10.1002/tea.21345

Šimunović, M., and Babarović, T. (2020). The role of parents' beliefs in students' motivation, achievement, and choices in the STEM domain: a review and directions for future research. *Soc. Psychol. Educ.* 23, 701–719. doi: 10.1007/s11218-020-09555-1

Sobiraj, S., Rigotti, T., Weseler, D., and Mohr, G. (2015). Masculinity ideology and psychological strain: considering men's social stressors in female-dominated occupations. *Psychol. Men Masc.* 16, 54–66. doi: 10.1037/a0035706

Stevens, T., Wang, K., Olivárez, A., and Hamman, D. (2007). Use of Self-perspectives and their Sources to Predict the Mathematics Enrollment Intentions of Girls and Boys. *Sex Roles* 56:351. doi: 10.1007/s11199-006-9180-2

Sullivan, J., Moss-Racusin, C. A., Lopez, M., and Williams, K. (2018). Backlash against gender stereotype-violating preschool children. *PLoS One* 13:e0195503. doi: 10.1371/journal.pone.0195503

Thompson, E. H., and Pleck, J. H. (1986). The structure of male role norms. *Am. Behav. Sci.* 29, 531–543.

Tomasetto, C., Mirisola, A., Galdi, S., and Cadinu, M. (2015). Parents' math–gender stereotypes, children's self-perception of ability, and children's appraisal of parents' evaluations in 6-year-olds. *Contemp. Educ. Psychol.* 42, 186–198. doi: 10.1016/j.cedpsych.2015.06.007

van der Vleuten, M., Jaspers, E., Maas, I., and van der Lippe, T. (2016). Boys' and girls' educational choices in secondary education. The role of gender ideology. *Educ. Stud.* 42, 181–200. doi: 10.1080/03055698.2016.1160821

Vandello, J. A., and Bosson, J. K. (2013). Hard won and easily lost: a review and synthesis of theory and research on precarious manhood. *Psychol. Men Masc.* 14, 101–113. doi: 10.1037/a0029826

Wang, M.-T. (2012). Educational and Career Interests in Math: a Longitudinal Examination of the Links Between Classroom Environment, Motivational Beliefs, and Interests. *Dev. Psychol.* 48, 1643–1657. doi: 10.1037/a0027247

Wang, M.-T. M.-T., Eccles, J. S., and Kenny, S. (2013). Not lack of ability but more choice: individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychol. Sci.* 24, 770–775. doi: 10.1177/0956797612458937

Wigfield, A., and Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemp. Educ. Psychol.* 25, 68–81. doi: 10.1006/ceps.1999.1015

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 February 2022
doi: 10.3389/fpsyg.2022.812483

Gender Differences in Self-Estimated Intelligence: Exploring the Male Hubris, Female Humility Problem

David Reilly[1*†], *David L. Neumann*[1,2†] *and Glenda Andrews*[1†]

[1] School of Applied Psychology, Griffith University, Southport, QLD, Australia, [2] Menzies Health Institute Queensland, Southport, QLD, Australia

OPEN ACCESS

Edited by:
Joyce Endendijk,
Utrecht University, Netherlands

Reviewed by:
Adrian Furnham,
University College London,
United Kingdom
Simmy Grover,
University College London,
United Kingdom
Irene Solbes Canales,
Complutense University of Madrid,
Spain

***Correspondence:**
David Reilly
davidreilly.com@gmail.com

†ORCID:
David Reilly
orcid.org/0000-0003-1227-9806
David L. Neumann
orcid.org/0000-0001-5400-462X
Glenda Andrews
orcid.org/0000-0001-8047-9151

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: 10 November 2021
Accepted: 13 January 2022
Published: 07 February 2022

Citation:
Reilly D, Neumann DL and Andrews G (2022) Gender Differences in Self-Estimated Intelligence: Exploring the Male Hubris, Female Humility Problem. Front. Psychol. 13:812483. doi: 10.3389/fpsyg.2022.812483

Despite evidence from cognitive psychology that men and women are equal in measured intelligence, gender differences in self-estimated intelligence (SEI) are widely reported with males providing systematically higher estimates than females. This has been termed the *male hubris, female humility* effect. The present study explored personality factors that might explain this. Participants (N = 228; 103 male, 125 female) provided self-estimates of their general IQ and for Gardner's multiple intelligences, before completing the Cattell Culture Fair IQ test as an objective measure of intelligence. They also completed the Bem Sex Role Inventory (BSRI) as a measure of sex-role identification, and measures of general and academic self-esteem. Both gender and sex-role differences were observed for SEI, with males and participants of both genders who scored high in masculinity offering higher self-estimates. By comparing estimated and observed IQ, we were able to rule out gender differences in overall accuracy but observed a pattern of systematic underestimation in females. An hierarchical multiple regression showed significant independent effects of gender, masculinity, and self-esteem. Mixed evidence was observed for gender differences in the estimation of multiple intelligences, though moderately sized sex-role differences were observed. The results offer a far more nuanced explanation for the male hubris, female humility effect that includes the contribution of sex role identification to individual and group differences.

Keywords: gender differences, self-estimated intelligence, self-esteem, sex-roles, sex differences, human intelligence, education

INTRODUCTION

"Such is the nature of men, that howsoever they may acknowledge many others to be more witty, or more eloquent, or more learned; yet they hardly believe there be many so wise as themselves."—Thomas Hobbes, English philosopher.

Intellectual self-image can be a powerful predictor of eventual educational achievement. How we see ourselves intellectually—either as smart, academically capable or possessing more mediocre abilities—can have a profound impact on academic engagement and motivation, the pursuit of intellectual endeavors, persistence in the face of adversity, and self-efficacy beliefs, and even performance on tests of intellectual ability. Psychologists and educators have known this for decades, ever since Rosenthal and Jacobson's (1968) classic *Pygmalion in the Classroom* study. In this study, the experimenters had students complete a bogus "Harvard intellectual assessment," and

then teacher expectations of individual students were experimentally manipulated by randomly assigning children to be labeled as either "gifted," ordinary, or below average. Longitudinal testing found that those in the experimental group exhibited significant growth relative to their peers in psychometrically measured IQ one year later. Such an example highlights not only the self-fulfilling prophecy of intellectual self-image (Greven et al., 2009), but also that it can be manipulated and shaped by environmental factors outside of our own control or even awareness. Though the study has at times been criticized on methodological grounds (c.f., Rosenthal, 1995; Snow, 1995), it spurred research into the benefits of teaching a "growth mindset," and that intelligence is malleable rather than being innately fixed at birth (Dweck, 2016; Yeager et al., 2019).

This brings us to a quite curious phenomenon frequently observed in psychological studies over the last few decades: that, when asked to provide an estimate of their intelligence, males frequently provide higher estimates than females. Indeed, this pattern of gender differences in self-estimated intelligence (SEI) is so universally found across different samples, ages, ethnicities and cultures that it has been termed the *male hubris, female humility* (MHFH) problem by Furnham et al. (2001). It remains so interesting because there is overwhelming consensus in cognitive psychology that males and females *do not differ* in general intelligence; gender differences are only found for specific cognitive abilities like verbal/visual-spatial tasks rather than psychometric intelligence (for a thorough review see Halpern et al., 2011). However our appraisals of our intellect contribute greatly to academic motivation (Dweck, 2002)—students who feel that they are less intellectually capable than their peers are less motivated (Kornilova, 2009). This is particularly so in stereotypically female underrepresented fields like science, technology, engineering and mathematics (Reilly and Hurem, in press). Intellectual self-image also guides course selection (such as the decision to pursue more advanced coursework in high school and college), as outlined by Eccles (2013) expectancy-value theory of achievement motivation: students select coursework that they expect they can reasonably master, and shy away from more challenging subjects if they believe they are not smart enough.

The male hubris/female humility effect has sparked much research across more than thirty countries (Freund and Kasten, 2012; Furnham, 2017). An earlier meta-analysis of studies by Syzmanowicz and Furnham (2011) found the effect to be robust with an average effect size of $d = 0.37$, which is a small to moderate effect size. So, if males and females do not differ in general intelligence but males provide higher estimates of their own intellectual prowess than females, what factors explain this discrepancy?

ACCURACY OF SELF-ESTIMATED INTELLIGENCE

Psychological research has investigated whether people are accurate judges of their intellectual ability generally (irrespective of gender). This arises from several strands of investigation: firstly, whether people are generally sound judges of their intellectual strengths and weaknesses, and secondly, whether there are cognitive biases that affect such evaluations. Furnham (2017) noted that an unresolved research question is whether males over-estimate their actual IQ, females under-estimate IQ, or indeed both, but writes "there are not enough good studies with both self-estimated and test-derived IQ to settle the argument," p. 110. This may be due to the relative ease with which self-estimates of intelligence may be obtained, but the greater difficulty, time, and expense needed to administer psychometrically valid intelligence tests. There are some examples where a proxy is used, such as a vocabulary test, to investigate the association between self-estimated IQ and intellect ($r = 0.25$, McCrae and Costa, 1985), while others choose to use a test of non-verbal reasoning like the Raven's Progressive Matrices ($r = 0.29$, von Stumm, 2014). In a review of studies comparing self-estimated and psychometrically assessed intelligence, Paulhus et al. (1998) note that in student subject pool samples, correlations rarely exceed $r = 0.30$ which is a moderately sized effect. They further note that somewhat larger correlations are found in studies that sample from the general population. To provide a frame of reference for evaluating this, self-reports of intelligence have roughly the same predictive validity and accuracy as the situational judgment tests (SJTs) that are widely employed in organizational psychology for predicting cognitive performance ($r = 0.29$ in a meta-analysis by McDaniel et al., 2007). People's impressions of their intellect are therefore grounded firmly in reality, but their accuracy is subject to distortion by cognitive biases.

Cognitive Biases

One such bias noted in the literature is the "above-average effect" (Alicke, 1985; Dunning et al., 1989; Kruger, 1999; Kruger and Dunning, 1999), which holds that for socially desirable traits such as competence and intellectual ability, there is a tendency for most people to see themselves as better than the average person. The implication of this, Kruger and Dunning (1999) argue, is that such overly favorable views of their abilities mean that a large proportion of the population is "unskilled and unaware of it," p. 1121. Such a claim stands in contrast to evidence on the general accuracy of self-estimates of intelligence reviewed above, though the number of studies empirically testing this with psychometrically valid IQ tests are few.

Another bias is the self-esteem bias (Felson, 1981), which is the tendency for people to evaluate themselves in a way that is consistent with their general self-esteem; someone who is high in self-esteem will tend to see themselves as brighter and more capable than someone lacking in self-esteem. While self-esteem is a normally distributed trait, there are frequently observed variations for different subgroups. Gender differences in general and academic self-esteem are well documented (Eccles et al., 1993; Gentile et al., 2009), with boys and men reporting higher general and academic self-esteem than girls and women. Syzmanowicz and Furnham (2011) raised this issue in their meta-analytic review as one possible explanation for the MHFH effect. However, they reported no correlation between self-estimated

intelligence and self-esteem, and it seems few studies have actually pursued this line of reasoning (Mirjalili et al., 2011).

Parental Beliefs, and Socio-Cultural Transmission of Gender Stereotypes

Environmental factors are also likely to contribute to a gender bias in self-estimated intelligence which may be an extension of existing socio-cultural gender stereotypes. Social motives (e.g., boastful pride for males or modesty for females) might explain self-estimates of intelligence. If so, when asked to estimate of *other people's* intelligence the MHFH effect should not still be present. In the original study by Hogan (1978) into self-estimates of intelligence, participants were also asked to provide an estimate of the intelligence of their mothers and fathers. Fathers were rated as more intelligent than mothers (Hogan, 1978), even though there were no gender differences in general intelligence in the community. The effect has been replicated numerous times (Beloff, 1992; Furnham and Rawles, 1995), but should be interpreted cautiously as it might reflect the systemic educational and occupational inequalities of the time (i.e., higher male educational advancement) rather than genuinely held beliefs that men are inherently "smarter."

Furnham and Gasson (1998) took a different approach, and instead asked parents to provide an estimation of the intelligence of their own children. Sons were rated as more intelligent than daughters (d = 0.67), and this effect has been replicated (Beloff, 1992; Furnham, 2000; Furnham et al., 2002a). Such a pattern of results suggests that environmental factors like gender stereotypes might contribute to the MHFH problem, rather than differential social desirability for intelligence between men and women. Parental beliefs may be a particularly important mechanism in the socialization of gender stereotypes, as parental educational expectations may influence a child's view of their own capabilities (Frome and Eccles, 1998; Jodl et al., 2001). Parental beliefs and expectations may inadvertently enhance or stifle a developing child's intellectual self-concept and self-efficacy beliefs: raising a child that feels either bright and capable even in the face of challenges (mastery orientation) or overwhelmed and incapable of more advanced intellectual achievement (learned helplessness). Numerous studies have demonstrated that parental beliefs about their children's intellectual abilities predict later educational achievement in adolescence and young adulthood (Jodl et al., 2001; Phillipson and Phillipson, 2007; Gunderson et al., 2012; Pinquart and Ebeling, 2019). This may be partly through direct transmission of parental beliefs and expectations, but also because parents can provide or withhold enriching cognitive experiences which can accelerate intellectual development outside of school.

Parents are but one element in a larger ecological system that contributes to intellectual development and intellectual self-image. This system includes the role of teachers and educators in shaping the intellectual self-image of children in their care (Jussim and Harber, 2005; Kollmayer et al., 2018), as well as differential treatment of boys and girls (particularly in gender-typed courses such as mathematics and science). Children's intellectual self-image is also shaped by media and popular culture (Solbes-Canales et al., 2020), which also plays a part in transmission of cultural gender stereotypes about intellectuality (Nosek et al., 2002; Storage et al., 2020).

Sex-Role Identification and Self-Estimated Intelligence

Another potential explanation for the MHFH effect may be the contribution of gendered personality traits, and sex-role identification. Bem (1981b) proposed *gender schema theory* as a cognitive account for the way that cultural prescriptions about masculinity and femininity become integrated into our self-concepts. These self-concepts forms internalized standards for regulating our own behavior, and also evaluating that of others through the lens of a gender schema. Now, while boys and girls typically differ in their early socialization experiences (Eccles et al., 1990; Lytton and Romney, 1991), there is also considerable individual variation in the degree to which one acquires stereotypically masculine and feminine personality traits, behaviors and interests- a process termed sex-typing (Kagan, 1964; Kohlberg and Ullian, 1974). The internalized gender schema of each individual differs and is the product of both biological and environmental factors that contribute to their sex-role identity (Tenenbaum and Leaper, 2002; Hines, 2011, 2015; Svedholm-Häkkinen et al., 2018). Highly sex-typed persons are motivated to keep their behavior and self-concept consistent with traditional gender norms of their biological sex (Maccoby, 1990; Martin and Ruble, 2004), and so implicit beliefs about gender and intellectuality could translate to higher estimates of intelligence by males and lower estimates by females. For many people their sex-role identification is veridical with their biological sex, but others are more flexible and incorporate a healthy blend of both masculine and feminine personality traits into their self-schema. Researchers have termed this psychological androgyny (Bem, 1984; Spence, 1984; Reilly, 2019), and it has been associated with greater psychological adaptability and less rigid gender schemas. Might sex-role identification act as a better predictor of self-estimated intelligence than the social category of gender?

There are several lines of reasoning that would support such an association. Firstly, as outlined above, it has been hypothesized that self-esteem makes a strong contribution to self-estimated intelligence. While gender differences in self-esteem are frequently reported (Gentile et al., 2009), numerous studies have documented a positive association between masculinity and self-esteem in both men and women (Whitley, 1983; Burnett et al., 1995). This, in turn, might drive higher self-estimates of intelligence. Secondly, there are links between sex-role identification and the development of cognitive ability. Nash's (1979) sex-role mediation hypothesis proposed that both masculine and feminine sex-roles contribute to cognitive development: masculinity predicts visual-spatial performance (Reilly and Neumann, 2013), while femininity predicts verbal and language abilities (Pajares and Valiante, 2001; McGeown et al., 2011; Reilly et al., 2016). Those higher in masculine and feminine traits may rate their abilities in those domains as higher, which may contribute to their overall impression of

intellectuality. Beyer and Bowden (1997) reported the tendency for women to underestimate their performance on stereotypically masculine tasks, but that this underestimation was not found for neutral or feminine tasks. Thirdly, for those with rigid gender schema, male boastfulness and female humility may temper their self-reports and over time shape their self-concept to reflecting implicit gender stereotypes.

Several studies have tested the contribution of sex-role identification to the MHFH effect. The first by Furnham et al. (1999) recruited a small number of subject pool participants, and had them complete the Personal Attributes Questionnaire (PAQ; Spence et al., 1974) which attempts to measure masculinity and femininity as personality traits. Results were inconclusive, though the study was underpowered. A second study by Rammstedt and Rammsayer (2002) recruited a larger sample size and instead used the Bem Sex-Role Inventory (BSRI; Bem, 1981a) which has greater psychometric validity (Choi et al., 2009). Subjects were asked both about their overall intelligence, as well as domain-specific multiple intelligences in line with Gardner's (1999) typologies. The authors found tentative support for sex-role effects in males, with those scoring higher in masculinity rated their mathematical-logical and general reasoning higher than lower-masculinity peers. However, the authors did not find sex-role effects for the females in their sample. Finally, a study by Storek and Furnham (2012) that recruited intellectually gifted MENSA members found a positive association between masculinity and self-estimated intelligence in both men and women. However, generalizability from such a highly-select sample is questionable. Furthermore, none of these studies included an actual measure psychometric IQ or of self-esteem to determine what role (if any) this played in the MHFH effect.

GENERAL INTELLIGENCE VERSUS MULTIPLE INTELLIGENCES

Experts on human intelligence have different views on the nature and structure of intelligence to those of the everyday man and woman. Intelligence is not a unitary construct (Neisser, 1979; Halpern, 2011), and comprises a large number of distinct abilities such as verbal intelligence, mathematical/logical intelligence, emotional intelligence, and so on. Sternberg et al. (1981) examined how lay conceptions of intelligence cluster around a different set of abilities to that of intelligence experts. Sternberg (2000, p. 3) argued that understanding these implicit or lay theories of intelligence was crucial, as "implicit theories of intelligence drive the way in which people perceive and evaluate their own intelligence and that of others." In reference to the present topic, while gender differences in overall SEI are widely documented, we might see different estimation patterns for certain abilities, such as those stereotypically regarded as masculine or male-dominated (mathematical/analytical, spatial), and those more readily associated with femininity or that are regarded as stronger in females (e.g., verbal and emotional intelligence).

One taxonomy for considering intelligence is Gardner's (1983, 1999) theory of multiple intelligences. Furnham (2000, 2001) first investigated whether the MHFH effect extended to Gardner's multiple intelligences, which has since been expanded to encompass seven to nine distinct clusters of abilities depending on the definitions used (Furnham et al., 2001, 2002a,b). Even though intelligence researchers may disagree on the psychometric validity of Gardner's multiple intelligences, student perceptions of them are important as they may guide course selection. Subjects are typically presented with a definition of each of Gardner's multiple intelligences, and asked to estimate their intelligence relative to others. These domains are: verbal or linguistic intelligence, logical or mathematical intelligence, spatial intelligence, musical intelligence, bodily-kinesthetic intelligence, interpersonal intelligence, intrapersonal intelligence, naturalistic intelligence, and existential/spiritual intelligence.

Research on self-estimations has revealed a complex and nuanced pattern: while gender differences were almost always found for estimates of general intelligence, they were not reliably found for all of Gardner's multiple intelligences. Moreover, cross-cultural differences are present. For example, Yuen and Furnham (2006) found that students in Hong Kong did not exhibit significant gender differences for verbal or interpersonal intelligence (stereotypically feminine) but did for all the remaining abilities. However, Furnham et al. (1999, Study 2) found significant gender differences with an English sample for only three of Gardner's domains: mathematical/logical, spatial and musical intelligence. A review by Furnham (2001) on several of Furnham and colleagues' studies noted that consistent gender differences were primarily found on stereotypically masculine intellectual abilities (mathematical/logical, and spatial), which Storek and Furnham (2012, 2014) subsequently referred to as domain-masculine intelligence (DMIQ). Furthermore, Storek and Furnham (2013) also found a moderately sized correlation between masculinity and self-estimates for DMIQ, $r = 0.26$, suggesting that there may be sex-role contributions to the effect.

When there are inconsistencies across studies and types of samples, the technique of meta-analysis provides a greater degree of confidence of the robustness of an effect than any single study alone. Syzmanowicz and Furnham (2011) conducted a meta-analysis on self-estimates of general intelligence and for three multiple intelligence domains, reporting moderately large gender differences favoring males for general intelligence, $d = 0.37$, mathematical/logical intelligence, $d = 0.44$, spatial intelligence, $d = 0.43$, and a much smaller difference for verbal intelligence, $d = 0.07$. However, none of the other forms of multiple intelligences were investigated. Moreover, further research is required to determine the extent of gender differences for other domains and to test potential moderators for the self-estimation effects.

THE PRESENT STUDY

Given the limitations outlined above with previous studies, we set out to explore potential factors that might explain the male-hubris, female humility (MHFH) effect. As Furnham (2017) remarked, there is a paucity of studies comparing

self-estimates to psychometrically valid IQ scores, and there are fewer still that include a measure of general self-esteem. Including a measure of sex-role identification would allow us to determine whether social category (male or female) or personality traits (masculinity/femininity) is a better predictor of SEI. Our investigation was primarily exploratory in nature rather than advocating for a particular theory, and this required us to perform additional statistical tests (e.g., that there might be gender differences in psychometric intelligence for our sample due to selection bias) in order to rule them out as alternate explanations. Also, though not the primary focus of the study, previous literature had identified associations between sex-role identification and self-esteem, as well as consistent gender differences. On this basis, it was reasoned that self-esteem might partly explain gender and sex-role differences in self-estimated intelligence (SEI).

The following hypotheses were made:

H1) Males will report higher SEI scores than females for general intelligence.
H2) High masculinity participants (i.e., masculine and androgynous groups) will report higher SEI scores than low masculinity ones (i.e., feminine and undifferentiated), regardless of gender.
H3) Males and high masculinity groups will report higher general self-esteem and academic self-esteem than females and low masculinity groups.
H4) There would be a significant positive correlation between SEI and psychometric intelligence, consistent with past studies (Paulhus et al., 1998).
H5) It is hypothesized that gender, masculinity, and general self-esteem will be associated with SEI, even after controlling for psychometrically measured intelligence.
H6) Masculinity scores would act as a statistical mediator of the relationship between gender and SEI scores.
H7) Gender and sex-role differences will also be found in self-estimates of multiple intelligences, following a similar pattern as observed with general intelligence.

MATERIALS AND METHODS

Participants

Two hundred and twenty-eight participants (103 male, 125 female) with a mean age of 22.62 (SD = 6.30, range = 18–47 years) were recruited from a university subject-pool of students completing a first-year research methods and statistics course. While the majority of these students were completing an undergraduate psychological science degree (53.7%), a large proportion were enrolled in exercise science or physiotherapy (30.7%), followed by health or biomedical sciences (7.4%) and occupational therapy (4%). Only 3% were studying another type of degree. This subject pool was chosen because it included psychology and non-psychology students in order to draw from a broader pool of sex-role categories. Most students were in their first trimester of university and participated prior to receiving their course grade. All participants provided informed consent to a research protocol approved by the Griffith University Human Research Ethics Committee (HREC).

Procedure

Participants were informed that they were participating in a study on the measurement of human intelligence, and the accuracy of self-estimates. They were provided with a booklet containing the self-estimated intelligence (SEI) measures, followed by the Cattell Cultural Fair IQ Test (CCFIT). Rest periods were provided between each subtest of the CCFIT test to minimize fatigue effects. Following test administration, participants completed surveys measuring self-esteem, sex-role identification, and general demographic information. The surveys were administered after the self-estimated intelligence survey and CCFIT, in order to minimize gender priming effects on SEI and test performance. Participants were tested in small groups (maximum three participants per session) so that compliance with instructions could be monitored and that survey items were read and considered before answering.

Measures

Self-Estimated Intelligence

Following the methodology of Furnham and Rawles (1995), participants were provided with a simple one page sheet from the booklet which explained in a brief paragraph that the distribution of intelligence in the general population followed a bell curve (see **Figure 1** for stimuli) that is normally distributed, with the average IQ score being 100 with a standard deviation of 15. The text of the paragraph was also read aloud by the experimenter to ensure that written instructions were followed. While the properties of the normal distribution were familiar to students in the statistics course, labeled framing anchors were also provided to aid in estimation. Participants were asked to use this scale to provide an estimate of their intelligence relative to other people, and to write this as a whole number.

Moderate Mild Below Average Above Gifted Exceptionally

Impairment Average Gifted

On a subsequent page of the booklet, participants read several paragraphs describing the research of Gardner's (1999) theory of multiple intelligences, which defined intelligence more broadly than would be typically assessed by an IQ test. Gardner subsequently revised his model of multiple intelligences to include a total of nine separate skills (Verbal and linguistic intelligence, Logical-mathematical intelligence, Spatial, Musical, Bodily-kinesthetic, Interpersonal, Intrapersonal, Naturalistic, and Existential/Spiritual Intelligence). Each skill was accompanied by a brief paragraph description that had been pilot tested for readability. An issue identified in pilot testing was that some participants completed the task extremely quickly with minimal variation in scores across domains. So that participants gave considered and deliberated responses, they were instructed to complete the task one definition at a time, and to record a response *only* after the experimenter had read the paragraph

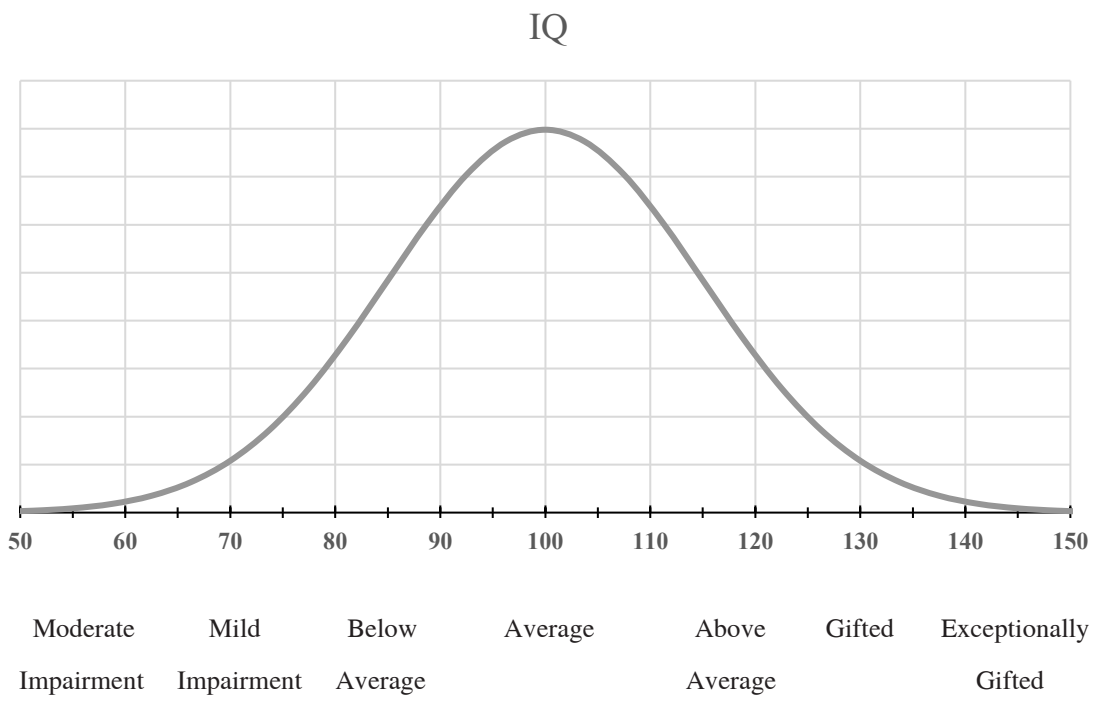


FIGURE 1 | Stimulus material used for self-estimation of intelligence.

aloud (on the pretense "some participants might come from a non-English background or have reading impairments such as dyslexia, and we want to make sure instructions are clearly understood"). This also ensured that participants had received the appropriate definition for each task, even if they elected not to read the presented material. The definition of existential/spiritual intelligence was phrased for inclusiveness so that it was clear to subjects that this may include but does not require religious practice. Participants responded by providing a numerical IQ score in the same format as for general intelligence.

Cattell Culture Fair Test of Intelligence (Cattell et al., 1973)

The CCFIT is a non-verbal measure of fluid intelligence (*gF*), designed specifically to be as free of culture and educational experiences as possible. Additionally, prior research confirmed no gender bias in the CCFIT with equivalent scores for males and females among adult high school graduates (Colom and García-López, 2002). The specific instrument employed was CCFIT Scale 3, Form A intended for use with adult participants. The CCFIT assessment requires inductive reasoning about perceptual patterns, and is comprised of four subtests (series completion, classification, matrices, conditions/typology). Each subtest is completed under strict timing conditions, with items of increasing level of difficulty such that less than 10% of subjects completed all items in the current sample. Although there was no penalty for guessing, two of the subtests require multiple correct responses for the item to be scored correctly. Individual responses were recorded on response sheets that were transcribed and then computer scored for accuracy of scoring. Reliability of the instrument for the current sample was high across the four subtests (Cronbach's $\alpha = 0.72$).

The instrument also provides appropriate norms tables to allow for conversion between raw scores and their equivalent IQ (centered around a mean of 100 with a standard deviation of 15), for direct comparability to SEI scores provided by participants. The CCFIT also shows strong convergent validity other tests of general intelligence such as the WAIS with $r = 0.72$ (Cattell et al., 1973), and loads highly against more recently revised intelligence scales (Carroll, 1993).

General Self-Esteem

Participants completed the Rosenberg (1965) General Self Esteem Scale, a brief 10 item rating scale that is widely used and demonstrates good psychometric reliability and validity (Sinclair et al., 2010). Participants recorded a response on a 4-point Likert-type scale (ranging from 1 = "*Strongly Agree*," to 4 = "*Strongly Disagree*"). Sample items include "On the whole, I am satisfied with myself" and "All in all, I am inclined to feel that I am a failure," with several items being reverse coded (Cronbach's $\alpha = 0.89$ for sample).

Academic Self-Esteem

There were two measures. Subjects completed a seven-item Academic Self-Esteem scale adapted for this present study from Johnson et al.'s (1983) Academic Self-Esteem subscale, and Bachman's (1970) Self-Concept of Ability Scale (SCAS). For comparability, subjects endorsed items on the same 4-point scale used for the Rosenberg GSES. Sample items include "I feel confident in my ability to complete university," and "I am not

doing as well at university as I would like to" with negatively worded items that were reverse coded. Subjects also completed the single item Rosenberg Academic Self-Esteem scale, which asks "How do you rate yourself in academic ability compared with those studying your degree" on a 4-point scale. The final response variable incorporated both measures of academic self-esteem, with high reliability (Cronbach's $\alpha = 0.87$).

Bem Sex-Role Inventory

The 30-item short form of the Bem Sex Role Inventory (BSRI; Bem, 1974, 1981a) was used as a measure of sex-role identification that construes masculinity and femininity as independent constructs on a continuous scale (Reilly, 2019). The BSRI includes 10 masculine, 10 feminine as well as 10 neutral and filler items so that the gendered nature of the instrument is not transparent. Traits are rated on a 7-point Likert scale (from "1 = Never or almost never true of me" to a midpoint of "4 = Occasionally true" and ending in "7 = Always or almost always true of me"). Separate masculinity and femininity scores were produced by averaging responses across each scale, resulting in a continuous score. Participants were also categorized on the basis of a median split of their masculinity ($Mdn = 4.60$) and femininity ($Mdn = 5.30$) scores, to one of four sex-role categories: masculine, feminine, androgynous (high masculinity and high femininity) and undifferentiated (low in both masculine and feminine personality traits). Internal consistency, as assessed by Cronbach's α, was high in the present sample (masculinity scale $\alpha = 0.81$, femininity scale $\alpha = 0.85$) and despite the passage of time since its inception the BSRI remains a valid measure of sex-role identification in modern samples (Choi et al., 2009). For a further review on the psychometric properties of sex-role measures, and why the BSRI remains valid today see Wood and Eagly (2015), and Eagly and Sczesny (2019).

RESULTS

We present first the sex-role classification for our sample, measured psychometric intelligence, self-estimated intelligence, general and academic self-esteem, followed by hypotheses testing.

Sex-Role Classification

The distribution of sex-role categories for participants appear in **Table 1**. As has been found in previous studies, the distribution of sex-role identification is not even in college-aged samples (e.g., feminine-scoring males and masculine-scoring females are underrepresented; Bem, 1981b). Also in line with past studies, independent samples *t*-tests showed that males were significantly higher in BSRI masculinity scores than females, $t(225) = 3.04$, $p = 0.003$, $d = 0.41$, and that females were significantly higher than males in BSRI femininity, $t(225) = -2.48$, $p = 0.014$, $d = -0.33$.

TABLE 1 | Distribution of sex-role categories in sample.

	Sex-role classification			
Gender	**Masculine**	**Feminine**	**Androgynous**	**Undifferentiated**
Males	29 (28.2%)	17 (16.5%)	34 (33.0%)	23 (22.3%)
Females	23 (18.5%)	33 (26.6%)	36 (29.0%)	32 (25.8%)

Cattell Culture Fair IQ Distribution

IQ scores for the sample were normally distributed (Shapiro–Wilks $p > 0.001$) with a mean of 111.19 ($SD = 14.21$). As might be expected from a university subject pool, a one-sample *t*-test showed our sample mean was significantly higher than that of the general population, $t(223) = 11.83$, $p < 0.001$, $d = 1.57$. Additionally, an independent samples *t*-test confirmed that males and females in our sample did not differ significantly in measured intelligence, $t(226) = 1.27$, $p = 0.206$. Any observed gender difference in SEI could not, therefore, be explained by apparent differences in actual intelligence between groups resulting from sampling error. Additionally, a 2 × (Gender) 4 × (Sex-Role Category) factorial ANOVA confirmed no sex-role differences in measured intelligence, nor any interaction, all $Fs < 2.61$, $p > 0.05$.

Self-Estimated IQ Distribution

The distribution of self-estimated intelligence scores in our sample was significantly negatively skewed (std. skewness = −2.19, $p = 0.028$), with a general tendency for participants to rate their intelligence as "above average," and a mean SEI of 107.55 ($SD = 10.98$). **Figure 2** presents a histogram of this distribution overlaid with the normal distribution of actual IQ scores in the general population ($M = 100$, $SD = 15$). Surprisingly though, approximately 19% of participants rated their intelligence as below average. This was somewhat unexpected as the "above average" effect had generally been regarded as robust—an issue we address further in the discussion. Additionally, there was a disproportionate number of

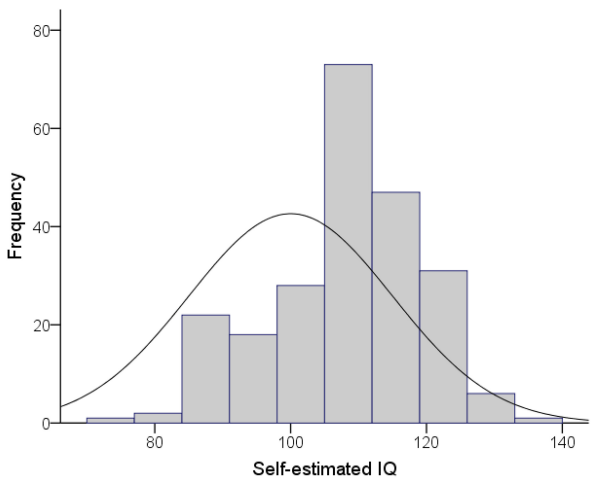


FIGURE 2 | Histogram showing the distribution of self-estimated intelligence in the present sample, alongside the normal distribution of IQ scores in the general population. Our sample was significantly negatively skewed with the bulk of scores shifted to the right of the normal curve.

females than males in this group, $\chi^2 = 24.08$, $p < 0.001$, with five males and 38 females rating their intelligence as below-average.

A 2 × (Gender) 4 × (Sex-Role Category) factorial ANOVA[1] was conducted on self-estimated IQ scores (see **Figure 3**). Although mild negative skewness was present (absolute standardized skewness = 2.23, $p < 0.05$), the ANOVA is robust against minor violations of normality when variances are equal (Field and Wilcox, 2017). The assumption of homogeneity of variance was met. As predicted by H1 there was a significant main effect of gender, $F(1, 219) = 30.79$, $p < 0.001$, $\eta^2 = 0.12$. Males ($M = 112.12$, $SD = 9.20$) reported significantly higher self-estimated IQ than females ($M = 103.66$, $SD = 10.88$), $t(225) = 5.55$, $p < 0.001$, $d = 0.74$, which equates to a difference of approximately 8.5 IQ points. There was also a significant main effect of sex-role category, $F(3, 219) = 7.23$, $p < 0.001$, $\eta^2 = 0.09$. A planned linear contrast compared the high masculinity participants (masculine + androgynous) to the low masculinity participants (feminine + undifferentiated). Consistent with H2, masculine and androgynous subjects gave higher self-estimates of IQ than feminine and undifferentiated, $t(225) = 4.65$, $p < 0.001$, $d = 0.62$. Both effects were medium in size. There was no significant interaction between gender and sex-role category.

General and Academic Self-Esteem

Next, we investigated individual differences in general and academic self-esteem, as these may make a contribution to perceptions of how intelligent our subjects perceived themselves to be. For the Rosenberg General Self-Esteem measure we conducted a 2 × (Gender) 4 × (Sex-Role Category) factorial ANOVA (see **Figure 4**). The data was normally distributed, and all assumptions were met. There was a significant main effect of gender, $F(1, 219) = 6.71$, $p = 0.010$, $\eta^2 = 0.03$, with males giving higher self-reports of general self-esteem than females ($d = 0.40$). Additionally there was a significant main effect of sex-role category, $F(3, 219) = 7.88$, $p < 0.001$, $\eta^2 = 0.10$, but no interaction between these terms. The effect of sex-role category was stronger than the social category of gender. In line with experimental hypotheses, a planned contrast confirmed that masculine and androgynous subjects reported higher general self-esteem scores than feminine and undifferentiated, $t(225) = 4.62$, $p < 0.001$, $d = 0.62$, which was a medium sized effect. Significant gender and sex-role differences indicate support for H3.

We repeated the factorial ANOVA for the academic self-esteem measure. As was the case with general self-esteem, males reported significantly higher academic self-esteem than females, $F(1, 219) = 15.01$, $p < 0.001$, $\eta^2 = 0.06$. A significant main effect of sex-role category, $F(3, 219) = 6.04$, $p = 0.001$, $\eta^2 = 0.08$, was also found. However, the interaction was not significant, and again the sex-role identification effect was slightly stronger than gender. The planned contrast demonstrated that participants with high masculinity (masculine and androgynous sex-roles) reported significantly higher academic self-esteem than participants with low masculinity (feminine and undifferentiated sex roles), $t(225) = 4.26$, $p < 0.001$, $d = 0.57$, which is a medium effect size.

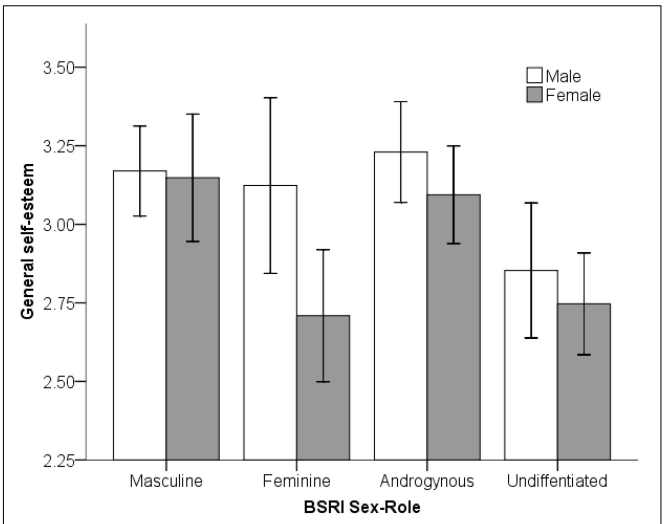


FIGURE 4 | Rosenberg General Self-Esteem scores across gender and sex-role categories (error bars represent the standard error of the mean).

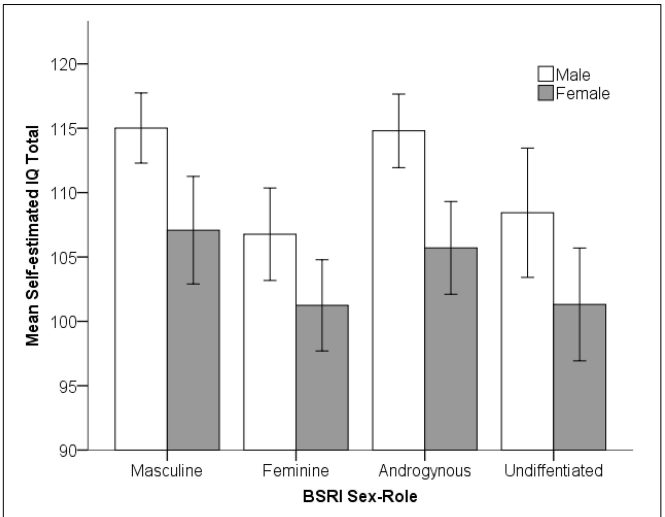


FIGURE 3 | Self-estimated IQ scores across sex-role categories, for males and females (error bars represent the standard error of the mean).

Bivariate Correlations

Bivariate correlations between all measures are reported in **Table 2**. Directions of correlations were consistent with previous literature, with gender and masculinity being significantly correlated with self-estimated IQ, both measures of self-esteem, and with IQ discrepancy scores (defined as self-estimated IQ—Cattel IQ). Additionally, self-estimated IQ was positively correlated with Cattell IQ scores, general self-esteem and academic self-esteem.

[1]A reflected log transformation was applied to the distribution and the analysis repeated, with no change in outcome. As the untransformed data was in a metric (IQ score) that was more meaningful, the untransformed data is reported. Additionally, the analysis was run with CCFIT as a covariate with no change in outcome.

TABLE 2 | Bivariate correlations between gender and sex-role measures, self-estimated intelligence, measured intelligence, general and academic self-esteem (N = 228).

Measure	1.	2.	3.	4.	5.	6.	7.	8.
1. Gender[a]	–	−0.21**	0.16*	−0.38***	−0.08	−0.20**	−0.20**	−0.27***
2. BSRI masculinity		–	0.02	0.34***	0.06	0.19**	0.37***	0.26***
3. BSRI femininity			–	−0.04	−0.07	0.04	0.11	−0.04
4. Self-estimated IQ				–	0.30***	0.44***	0.28***	0.45***
5. Cattell IQ					–	−0.72***	−0.02	0.08
6. IQ Discrepancy						–	0.22**	0.25***
7. Rosenberg Self-Esteem							–	0.54***
8. Academic Self-Esteem								–

*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.
[a]*Dummy coded variable; 0 = male, 1 = female.*

Predictors of Gender Differences in Self-Estimated Intelligence

Next, we set out to explore possible explanations for the male hubris, female humility effect. In the sample, the correlation between SEI and measured intelligence was just at the cusp of being medium in strength, $r(228) = 0.30$, $p < 0.001$, and the scatterplot confirmed it was linear in nature. This is consistent with past research that finds people are generally sound judges of their intelligence.

One possible explanation for the MHFH effect might be that males and females greatly differ in the *accuracy* of their judgments of self-estimated intelligence though. To rule out this explanation, we examined the bivariate correlation between SEI and measured intelligence for males and females separately (see **Figure 5**). The correlation between SEI and measured intelligence was slightly higher for males, $r(103) = 0.33$, $p < 0.001$, than for females, $r(124) = 0.26$, $p = 0.004$, but again, both fell in the small to medium range of effect sizes and any difference most likely reflects sampling error. To confirm this, Fisher's r-to-z transformation was applied to assess the significance of the difference between the two correlation coefficients r_{male} and r_{female}, $z_{dif} = 0.57$, $p = 0.284$ (1-tailed), indicating no difference. Thus, we were able to rule out the possibility of differences in accuracy between males and females as an explanation for the male hubris, female humility problem. As can be seen in **Figure 5** though, visual inspection does suggest a tendency for gender differences in direction, with more blue scores above the regression line and more green scores below.

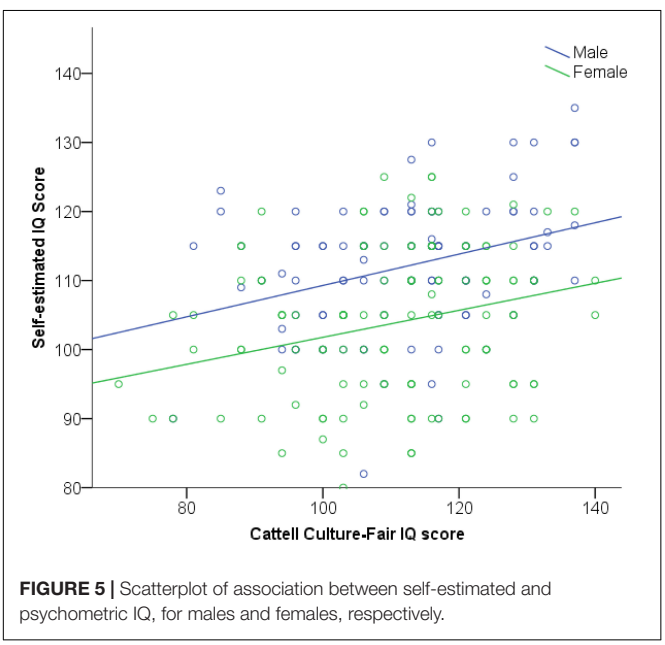


FIGURE 5 | Scatterplot of association between self-estimated and psychometric IQ, for males and females, respectively.

Another plausible explanation for gender differences in SEI might be the contribution of self-esteem. Reported in **Table 2**, there was a moderate positive correlation between self-estimated intelligence and general self-esteem scores. However, it is also plausible that having a high intellect also makes a positive contribution to one's general self-esteem, so we tested whether the correlation between SEI and general self-esteem remained significant after controlling for psychometric IQ. The positive correlation between SEI and Rosenberg General Self Esteem with CCFIT scores partialed out was still statistically significant, $r = 0.30$, $p < 0.001$, and of moderate strength (i.e., general self-esteem was associated with self-estimates of intelligence). As might be expected, the correlation between SEI and academic self-esteem was somewhat stronger, $r = 0.45$, though this is likely to be a bidirectional relationship.

To explore the joint effects of the social category of gender, sex-role identification, and general self-esteem, a hierarchical multiple regression was conducted on self-estimated intelligence scores (see **Table 3**). Psychometric IQ scores were entered at Step 1 in order to control for individual differences in actual intelligence, $F_{chg}(1,223) = 22.71$, $p < 0.001$, explaining approximately 9% of the variance in SEI. Next in Step 2, gender was entered in conjunction with sex-role identification (BSRI masculinity and femininity scores). Although only gender and masculinity were hypothesized to make a significant contribution to SEI scores, femininity was included to consider the possibility it also made a significant contribution. Together these factors resulted in an increased model fit, $F_{chg}(3,220) = 20.76$, $p < 0.001$, explaining an additional 20% of variance in the dependent variable of SEI. Both gender and masculinity scores were significant predictors. Finally at Step 3, General Self-Esteem scores were entered to test the hypothesis that self-esteem may still be a contributing factor. This resulted in a small increase in model fit, $F_{chg}(1,219) = 4.39$, $p < 0.001$. The final model was statistically significant, $F(5, 219) = 19.36$, $p < 0.001$, accounting

TABLE 3 | Hierarchical multiple regression of self-estimated intelligence scores (N = 228).

Variable		β	*t*	*p-value*	sr^2	*R*	R^2
Step 1						30	0.09
	Cattell IQ	0.30	4.70	<0.001***	0.09		
Step 2						0.54	0.29
	Cattell IQ	0.26	4.57	<0.001***	0.06		
	Gender (0 = male)	−0.31	−5.33	<0.001***	0.10		
	Masculinity	0.28	4.89	<0.001***	0.06		
	Femininity	0.02	0.39	0.700	0.00		
Step 3						0.55	0.31
	Cattell IQ	0.26	4.72	<0.001***	0.07		
	Gender (0 = male)	−0.29	−4.95	<0.001***	0.08		
	Masculinity	0.23	3.80	0.001**	0.05		
	Femininity	0.01	0.11	0.913	0.00		
	General Self-Esteem	0.13	2.19	0.030*	0.02		

*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

for 31.7% of the variance in individual self-estimates of intelligence. As can be seen from the table, even after controlling for individual differences in measured intelligence (β = 0.27), the three hypothesized predictors of gender, masculinity and general self-esteem made significant and unique contributions. Gender was the strongest predictor, followed by measured intelligence, masculinity, and finally a smaller contribution of general self-esteem which had considerable overlap with the other predictors.

Statistical Mediation

We next examined whether masculine sex-role identification (masculinity score as a continuous variable) acted as a statistical mediatior in the relationship between gender and SEI scores. Baron and Kenny (1986) proposed three criteria for establishing statistical mediation. Firstly, the predictor (gender) should predict the dependent variable (SEI). Secondly, the predictor must be correlated with the proposed mediator variable (masculinity, shown as Path A). Thirdly the mediator must correlate with the dependent variable (SEI) even after controlling for the contribution of the predictor (shown as Path B). The Sobel test of statistical mediation was significant, Sobel $z = -2.55$, $p = 0.010$, and calculation of the bootstrapped estimate of the indirect effect showed that it differed significantly from zero (95% CI = −2.26 to −0.41), following the bootstrapping criteria outlined in Preacher and Hayes (2004). As the mediation effect was significant, we then tested whether the relationship was fully or only partially mediated (Baron and Kenny, 1986). In a full mediation model, the association between predictor and dependent variable will no longer be statistically significant after controlling for the mediator (i.e., all of the effect of the predictor acts indirectly through the mediator, and does not make a direct contribution). This relationship is represented by Path C in **Figure 6**. Though diminished, the beta weight remained statistically significant, indicating that the relationship was only a partial mediation. Though acting indirectly through masculine sex-role identification, there was still a direct contribution of gender to SEI scores.

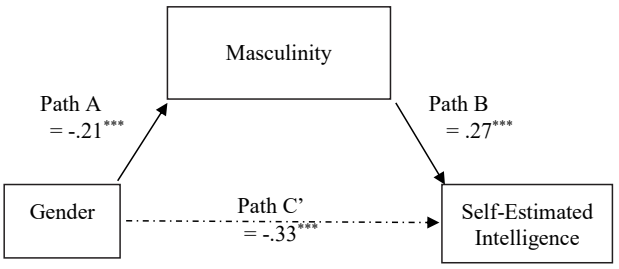


FIGURE 6 | Indirect effect of gender on SEI, with masculine sex-roles acting as a mediator on self-estimated intelligence. Path C' represents the direct effect of gender after controlling for the mediator. *** $p < 0.001$.

Having identified in the multiple regression analysis that biological sex made a slightly stronger contribution to SEI than measured intelligence, sex-role identification, and general self-esteem, we sought to quantify how large the discrepancy between self-estimates and measured intelligence was. A composite variable representing the difference between self-estimated and measured intelligence was created, with positive values indicating higher SEI than measured intelligence. An independent samples t-test on IQ discrepancy scores confirmed a significant gender difference, $t(225) = 3.04$, $p = 0.003$, $d = 0.40$. Visual inspection of the discrepancy scores showed that *on average*, males in our sample demonstrated fairly sound judgment in appraising their intelligence ($M = -0.35$, $SD = 13.61$), but that there was also wide variability with some males greatly overestimating their intelligence and some males underestimating (range = −27 to + 38 IQ points). However, females systematically *undervalued* their intellectual capabilities by over six IQ points ($M = -6.34$, $SD = 15.83$), and for those female participants that did offer inflated self-estimates, these were much smaller in size (range = −41 to + 25 IQ points). Only the female discrepancy scores differed significantly from zero however ($p < 0.001$).

Next, a 2 × (Gender) 4 × (Sex-Role Category) factorial MANOVA was performed on the nine self-estimates of Gardner's multiple intelligences. As the cell size differed across sex-role category and Box's M was significant ($p < 0.001$), Pillai's trace was selected as the more conservative estimate. Assumptions of normality and homogeneity of variance were met. In line with previous research, there was a significant multivariate effect of biological sex, $F(9, 212) = 7.02$, $p < 0.001$, $\eta^2 = 0.23$, which is a medium to large effect. There was also a significant multivariate effect of sex-role identification, $F(27, 642) = 2.22$, $p < 0.001$, $\eta^2 = 0.09$, though there was no significant interaction, $F(27, 642) = 1.02$, $p = 0.437$. As the overall multivariate effects were significant and of non-trivial size, this justified examination of univariate effects without a need to apply a Bonferroni correction (c.f., Huberty and Morris, 1989). For ease of comparison, sex and sex-role differences are reported separately in **Tables 4**, **5**, respectively. Five of the nine multiple intelligence domains showed significant differences between males and females, with effect sizes ranging from small to large.

Table 5 presents sex-role differences across the nine multiple intelligence domains. Although gender differences were not

TABLE 4 | Gender differences on self-estimated multiple intelligences.

Domain	Male	Female	$F_{(1,220)}$	*p*-value	*d*
1. Verbal	106.45 (12.87)	107.07 (11.65)	0.73	0.395	−0.05
2. Logical-Mathematical	108.39 (16.68)	98.66 (13.43)	18.36	<0.001***	0.64
3. Spatial	109.80 (12.51)	98.54 (11.93)	40.79	<0.001***	0.92
4. Musical	102.64 (18.11)	99.50 (14.72)	0.64	0.426	0.19
5. Bodily-kinesthetic	112.57 (14.26)	106.47 (14.74)	7.54	0.007**	0.42
6. Interpersonal	112.69 (12.98)	112.86 (11.72)	0.20	0.654	−0.01
7. Intrapersonal	110.61 (12.63)	109.36 (12.79)	0.11	0.742	0.09
8. Naturalistic	104.43 (11.88)	99.10 (11.06)	10.36	0.001**	0.46
9. Existential/spiritual	108.72 (16.92)	102.94 (12.84)	6.85	0.009**	0.39

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE 5 | Sex-role differences in self-estimated multiple intelligences.

Domain	Masc.	Fem.	Andr.	Undif.	F-ratio	Planned Contrast
1. Verbal	110.35 (1.61)	104.51 (1.81)	106.74 (1.51)	104.54 (1.66)	2.78*	$t(226) = 2.44$, $p = 0.015$, $d = 0.33$
2. Logical-Mathematical	106.66 (1.98)	98.41 (2.22)	104.38 (1.85)	103.23 (2.04)	2.68*	$t(226) = 2.32$, $p = 0.021$, $d = 0.32$
3. Spatial	105.71 (1.59)	100.34 (1.79)	107.01 (1.49)	101.83 (1.64)	3.70*	$t(226) = 3.22$, $p = 0.001$, $d = 0.43$
4. Musical	102.98 (2.12)	95.86 (2.38)	105.23 (1.99)	97.40 (2.18)	4.25**	$t(226) = 3.44$, $p = 0.001$, $d = 0.46$
5. Bodily-kinesthetic	112.18 (1.86)	106.22 (2.10)	114.00 (1.75)	103.96 (1.92)	6.47***	$t(226) = 4.18$, $p < 0.001$, $d = 0.56$
6. Interpersonal	113.82 (1.53)	112.36 (1.72)	117.87 (1.44)	105.83 (1.58)	10.82***	$t(226) = 4.30$, $p < 0.001$, $d = 0.57$
7. Intrapersonal	110.48 (1.66)	107.58 (1.87)	113.88 (1.56)	106.44 (1.71)	4.09**	$t(226) = 3.03$, $p = 0.003$, $d = 0.40$
8. Naturalistic	101.83 (1.50)	100.49 (1.69)	104.20 (1.41)	99.12 (1.55)	2.15*	$t(226) = 2.09$, $p = 0.038$, $d = 0.27$
9. Existential/spiritual	106.80 (1.93)	104.23 (2.17)	110.47 (1.81)	100.57 (1.99)	4.79**	$t(226) = 3.15$, $p = 0.002$, $d = 0.42$

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

present for every domain (**Table 4**), there were significant sex-role differences for each of the domains. Accordingly, a planned linear contrast was conducted comparing the high masculinity groups (masculine + androgynous) with the low masculinity groups. Masculine persons reported significantly higher self-estimates of multiple intelligences, with effect sizes ranging from quite small to medium in size.

DISCUSSION

The goal of the present study was to investigate psychological factors that contribute to the widely observed male hubris, female humility effect. These include baseline psychometric intelligence, general self-esteem, and sex-role identification (masculine and feminine personality traits). By including a suitable measure of psychometric intelligence, we were also able to rule out certain explanations for the MHFH effect, such as genuine differences in measured intelligence between the males and females recruited due to sampling bias, as well as discrepancies in the accuracy of self-estimated intelligence in one or both genders. What we found was a more nuanced picture that gives support to a multifactorial model for explaining gender differences in self-estimated intelligence. Furthermore, we found sex and sex-role differences for some but not all multiple intelligence domains, consistent with cultural gender stereotypes about certain intellectual domains. We first review support for the experimental hypotheses and then discuss the important social and educational implications of this pattern of results.

Hypothesis 1 was supported, with males reporting higher SEI scores than females as found in previous studies (Syzmanowicz and Furnham, 2011). Our observed effect (d = 0.74) was somewhat larger than that typically reported but in line with some studies reporting quite large gender differences (Bennett, 1996; Zhang and Gong, 2001). It may be driven in part by the sex-role composition of our sample as the planned contrast confirmed a similarly large sex-role effect (d = 0.62) between high and low masculinity subjects which supported Hypothesis 2. A previous study by Szymanowicz and Furnham (2013) also found sex-role differences, with masculinity predicting self-estimates of general intelligence and multiple intelligences consistent with our study.

However, inspection of the distribution of SEI scores highlights some key differences to that typically described in SEI studies. The better-than-average effect is widely regarded as a truism in the literature (Alicke et al., 1995) for it has been so widely reported (for a review see Kruger and Dunning, 1999). Paulhus et al. (1998) claim that "rarely do people rate themselves as "below average," p. 526. It is difficult to determine with any certainty how valid such claims are, however, as despite the large number of studies on self-estimated intelligence histograms are rarely presented. But on inspecting the histogram in **Figure 2**, there were an inordinately high number of participants that saw themselves as having below average intelligence. The stimulus material provided an anchoring frame of 100 as

"average intelligence," following the methodology of Furnham and colleagues. So it is unclear whether this results from a fundamental difference in the composition of our observed sample, or is instead a Hawthorne effect from the knowledge that subjects would soon complete a psychometrically valid IQ test. As noted, a number of studies use a proxy test like the Raven's Progressive Matrices, or the Wonderlic Personnel Test, but these stimulus materials do not explicitly identify them as IQ tests; thus they may be less imbued with power in the minds of our participants as an "actual IQ test." Might this have tempered somewhat the tendency to give inflated self-estimates for social desirability reasons, and altered behavior accordingly? Alternately, differences in the characteristics of our sample may be responsible, given the percentage of students from non-psychology faculties and broader diversity of sex-roles. Only replications of the study will be able to shed more light on this matter, but we strongly recommend that future studies report information on the distribution of SEI scores and the number of subjects who rate themselves as below average. It is also possible that it may be tapping into personality traits such as honesty-humility, as found by previous studies (Kajonius, 2014).

Consistent with previous studies, we also found predicted gender differences in general self-esteem and of comparable size (Gentile et al., 2009). Additionally, masculine/agentic personality traits appear to confer benefits for overall self-esteem as well as academic self-esteem (Whitley, 1983; Hirschy and Morris, 2002), which is important as this hypothesis has rarely been examined in recent years and may have been subject to shift as gender-norms change. Observed correlations between masculinity and general self-esteem (r = 0.26) are of similar strength to those reported in other studies (Hirschy and Morris, 2002). Thus, both sex and sex-role identification contribute jointly to self-esteem, affirming Hypothesis 3. To our knowledge, few researchers have empirically tested the contribution of general self-esteem to self-estimates of general intelligence. Only a single study, by Mirjalili et al. (2011) could be located, finding a correlation between Rosenberg Self-Esteem scores and SEI of $r = 0.32$ in a sample of Iranian boys and girls. Gender differences in general self-esteem though may in turn contribute to SEI scores, though causation cannot be established with a correlational design.

Subjects in our study were also reasonably astute judges of their own intellectuality, with a moderately sized positive correlation between SEI and psychometric IQ, consistent with Hypothesis 4. The observed effect size was comparable to previous studies (Paulhus et al., 1998), and one possibility that we were able to rule out as an explanation for the MHFH was that one or both genders held "completely unrealistic" views of their abilities as might be suggested from the label of male hubris or female humility. Instead, what emerged from the results of the multiple regression analysis was a multifactorial explanation for the MHFH effect. Affirming Hypothesis 5, there were significant independent contributions of biological sex, masculinity, and general self-esteem on self-estimates even after controlling for the contribution of psychometrically measured intelligence. Furthermore, the association between biological sex and SEI was statistically mediated by masculinity.

Additional investigation of discrepancy scores showed that males were fairly close in their self-estimates (but keep in mind this may have been tempered by the knowledge their estimates would be compared against their pending IQ test scores). However, there was still a tendency in females to underestimate their intelligence by an average of 6.32 IQ points (or almost half a standard deviation). This, when combined with the self-esteem and sex-role mediation effects, offers a broader explanation for the MHFH effect—we have a more nuanced picture than rather than simply the social category of male and female determining self-estimates.

While there was firm evidence for gender and sex-role differences in self-estimates of "global" intellectual ability for our sample, this did not extend to all multiple intelligences (Hypothesis 7). Consistent with Syzmanowicz and Furnham's (2011) meta-analysis there were gender differences for logical-mathematical as well as spatial intelligence which are stereotypically regarded as masculine and which the authors termed domain-masculine IQ (DMIQ). Across the stereotypically feminine intelligence domains of verbal, interpersonal and intrapersonal (collectively regarded by laypersons as "emotional intelligence") we did not see evidence of male hubris. However, despite being relatively small in magnitude, we did observe significantly higher male estimates for bodily-kinesthetic, naturalistic and existential/spiritual intelligence. In addition, there was firm support for sex-role differences in multiple intelligences, with masculine and androgynous subjects reporting higher estimates than feminine and undifferentiated. This also replicates results found in a previous study by Szymanowicz and Furnham (2013). Somewhat surprisingly however, these effects were observed even on stereotypically feminine intellectual domains, though again, the effect sizes were small-to-medium. The previous study failed to find significant sex-role differences on emotional intelligence (represented by inter- and intra-personal intelligence in the current study).

Social and Educational Implications

Though widely observed, for decades researchers have struggled to identify and understand the psychological factors contributing to the male hubris, female humility effect. While a mild self-enhancing bias may be protective and to some degree self-fulfilling, the psychological consequences of inaccurately calibrated estimates of intellectual ability can also be damaging. Unrealistically inflated estimates may set students up for future discouragement and failure if their reach exceeds their grasp: in the United States, college completion rates have been steadily declining and disproportionately affect males (Bound et al., 2010), while the percentage of Ph.D. students who start but do not complete a Ph.D. exceeds 50% (Most, 2008). Perhaps even more problematic though is the effect of underestimation on achievement motivation, course selection and educational aspirations (Eccles, 2013): if you tell yourself that you can't, then you're right—you won't. Systemic differences in self-estimated intelligence for an entire social class (women), has serious social, educational and financial implications that cannot be understated. Furthermore, we found evidence that

self-esteem and sex-role identification are also predictors of female hubris—and one must be mindful that there were also males in our sample who saw themselves as below average. Educators should also be reminded of the effect that praise and encouragement can have for students who underestimate their abilities, as well as the effect that negative gender stereotypes and implicit bias can have on impressionable young students. While disingenuous or inflated praise can sometimes backfire ("the praise paradox," Brummelman et al., 2014), as educators we should also be mindful that students might not always recognize their full potential. Dweck (2002) notes that strategic praise can increase academic motivation in such students (particularly when paired with a growth mindset), which also has a follow-on effect with course selection (Eccles, 2013).

Parental beliefs about differential intellectuality of sons and daughters reflect larger cultural beliefs that implicitly associate men and masculinity with intelligence (Nosek et al., 2002). Research on intelligence is unequivocal that men and women do not differ in objectively measured intelligence (Neisser et al., 1996), and endorsement of explicit beliefs that one gender is superior is quite rare (Swim, 1994). However, implicit beliefs differ, as do parental estimates of the intelligence of sons and daughters (Beloff, 1992; Furnham and Rawles, 1995). Rigid adherence to gender-roles, particularly in the educational context is problematic. Socio-cultural transmission of gender stereotypes (through parents, teachers, peers and media) may be in part fueling this phenomenon (Kollmayer et al., 2018), but there's also individual variability in the extent to which these are internalized. Masculine personality traits appear to be a protective factor, as well as predicting general self-esteem.

Limitations and Directions for Future Research

A previous study by Kajonius (2014) found that the personality trait of honesty-humility also predicted self-estimated intelligence, and that there are likely social desirability factors at play. Additionally, it is plausible that some participants may have had knowledge of either the male-hubris/female humility effect, or that of the Dunning-Krueger effect. There is also the possibility that students' self-perceptions of intellectual ability were shaped by being in an environment where they are provided feedback on assignments and examinations. While the ethnicity of students was not recorded, having an ethnically diverse sample which included international students might introduce the possibility of stereotype threat effects (Steele and Aronson, 1995), where students internalize negative stereotypes about particular ethnic groups.

As acknowledged above, we are also uncertain whether the number of people self-estimating themselves as below-average intelligence is a Hawthorne effect due to their knowledge that we would measure psychometric intelligence, or instead a difference in the composition of our sample. This is a point of difference from most previous studies on self-estimation of intelligence, as most previous studies have not coadministered an intelligence test. By their nature, university students are a self-selected sample and in Australia, go through the bottleneck of meeting certain educational achievement requirements. Could it be that lowered admission requirements and alternate entry pathways resulted in a markedly different sample to that found in typical American college subject pools? Or might the below-average effect observed here have been overlooked in previous studies? Further research is needed to explore this issue, and we advocate for coadministering measures of self-esteem and sex-role identification. It is hoped that further research will elucidate whether it is biological sex or psychological gender that better explains gendered patterns of self-estimated intelligence.

CONCLUSION

Possible explantions for the widely observed gender differences in self-estimated intelligence were investigated, which has also been termed the male-hubris female humility effect. We found that the issue is complex and nuanced, with no single cause emerging but rather that there were a number of contributing factors. Firstly, sex-role identification makes a significant contribution to intellectual self-image, with masculine/agentic personality traits leading to higher self-estimates. Secondly, we found a significant and independent effect of self-esteem to self-estimated intelligence. As females in our sample reported lower general self-esteem in line with the trend identified in the literature, this may be a strong factor underlying the male-hubris female humility effect. However further research is needed to elucidate the risk factors that identify patterns of over-/under-estimation of intelligence.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human Research Ethics Committee, Griffith University. The patients/participants provided their online informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DR, DN, and GA contributed to the conception and design of the study and contributed equally to the interpretation. DR completed the recruitment of participants and data analysis. All authors contributed to manuscript revision, read and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank Adrian Furnham for supplying a copy of the self-estimated IQ stimuli for replication purposes.

REFERENCES

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *J. Pers. Soc. Psychol.* 49, 1621–1630. doi: 10.1037/0022-3514.49.6.1621

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., and Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *J. Pers. Soc. Psychol.* 68, 804–825. doi: 10.1037/0022-3514.68.5.804

Bachman, A. M. (1970). The relationship between a seventh-grade pupil's academic self-concept and achievement in mathematics. *J. Res. Math. Educ.* 1, 173–179. doi: 10.2307/748335

Baron, R. M., and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173

Beloff, H. (1992). Mother, father and me: our IQ. *Psychologist* 5, 309–311.

Bem, S. L. (1974). The measurement of psychological androgyny. *J. Consult. Clin. Psychol.* 42, 155–162. doi: 10.1037/h0036215

Bem, S. L. (1981b). Gender schema theory: a cognitive account of sex typing. *Psychol. Rev.* 88, 354–364. doi: 10.1037/0033-295X.88.4.354

Bem, S. L. (1981a). *BEM Sex Role Inventory Manual: Manual, Instrument and Scoring Guide*. Mind Garden.

Bem, S. L. (1984). "Androgyny and gender schema theory: a conceptual and empirical integration," in *Nebraska Symposium on Motivation: Psychology and Gender*, ed. T. B. Sonderegger (Lincoln, NE: University of Nebraska Press), 179–226.

Bennett, M. (1996). Men's and women's self-estimates of intelligence. *J. Soc. Psychol.* 136, 411–412. doi: 10.1080/00224545.1996.9714021

Beyer, S., and Bowden, E. M. (1997). Gender differences in self-perceptions: convergent evidence from three measures of accuracy and bias. *Pers. Soc. Psychol. Bull.* 23, 157–172. doi: 10.1177/0146167297232005

Bound, J., Lovenheim, M. F., and Turner, S. (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *Am. Econ. J. Appl. Econ.* 2, 129–157.

Brummelman, E., Thomaes, S., Orobio de Castro, B., Overbeek, G., and Bushman, B. J. (2014). "That's not just beautiful—that's incredibly beautiful!" The adverse impact of inflated praise on children with low self-esteem. *Psychol. Sci.* 25, 728–735.

Burnett, J. W., Anderson, W. P., and Heppner, P. P. (1995). Gender roles and self-esteem: a consideration of environmental factors. *J. Couns. Dev.* 73, 323–326. doi: 10.1002/j.1556-6676.1995.tb01757.x

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge: Cambridge University Press.

Cattell, R. B., Krug, S. E., and Barton, K. (1973). *Technical Supplement for the Culture Fair Intelligence Tests, Scales 2 and 3*. Champaign, IL: Institute for Personality and Ability Testing.

Choi, N., Fuqua, D. R., and Newman, J. L. (2009). Exploratory and confirmatory studies of the structure of the Bem Sex Role Inventory Short Form with two divergent samples. *Educ. Psychol. Meas.* 69, 696–705. doi: 10.1177/0013164409332218

Colom, R., and García-López, O. (2002). Sex differences in fluid intelligence among high school graduates. *Pers. Individ. Differ.* 32, 445–451. doi: 10.1016/S0191-8869(01)00040-X

Dunning, D., Meyerowitz, J. A., and Holzberg, A. D. (1989). Ambiguity and self-evaluation: the role of idiosyncratic trait definitions in self-serving assessments of ability. *J. Pers. Soc. Psychol.* 57, 1082–1090. doi: 10.1037/0022-3514.57.6.1082

Dweck, C. (2016). What having a "growth mindset" actually means. *Harv. Bus. Rev.* 13, 213–226.

Dweck, C. S. (2002). "Messages that motivate: how praise molds students' beliefs, motivation, and performance (in surprising ways)," in *Improving Academic Achievement*, ed. J. Aronson (Amsterdam: Elsevier), 37–60.

Eagly, A. H., and Sczesny, S. (2019). Editorial: gender roles in the future? Theoretical foundations and future research directions [Editorial]. *Front. Psychol.* 10:1965. doi: 10.3389/fpsyg.2019.01965

Eccles, J. S. (2013). "Gender and achievement choices," in *Societal Contexts of Child Development: Pathways of Influence and Implications for Practice and Policy*, eds E. T. Gershoff, R. S. Mistry, and D. Crosby (New York, NY: Oxford University Press), 19–34. doi: 10.1093/acprof:oso/9780199943913.003.0002

Eccles, J. S., Jacobs, J. E., and Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *J. Soc. Issues* 46, 183–201. doi: 10.1111/j.1540-4560.1990.tb01929.x

Eccles, J. S., Wigfield, A., Harold, R. D., and Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Dev.* 64, 830–847. doi: 10.2307/1131221

Felson, R. B. (1981). Ambiguity and bias in the self-concept. *Soc. Psychol. Q.* 44, 64–69.

Field, A. P., and Wilcox, R. R. (2017). Robust statistical methods: a primer for clinical psychology and experimental psychopathology researchers. *Behav. Res. Ther.* 98, 19–38. doi: 10.1016/j.brat.2017.05.013

Freund, P. A., and Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychol. Bull.* 138, 296–321. doi: 10.1037/a0026556

Frome, P. M., and Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *J. Pers. Soc. Psychol.* 74, 435–452. doi: 10.1037/0022-3514.74.2.435

Furnham, A. (2000). Parents' estimates of their own and their children's multiple intelligences. *Br. J. Dev. Psychol.* 18, 583–594. doi: 10.1348/026151000165869

Furnham, A. (2001). Self-estimates of intelligence: culture and gender difference in self and other estimates of both general (g) and multiple intelligences. *Pers. Individ. Differ.* 31, 1381–1405. doi: 10.1016/S0191-8869(00)00232-4

Furnham, A. (2017). Sex differences in self-estimated intelligence, competiveness and risk-taking. *Mankind Q.* 58, 109–111. doi: 10.46469/mq.2017.58.1.9

Furnham, A., Clark, K., and Bailey, K. (1999). Sex differences in estimates of multiple intelligences. *Eur. J. Pers.* 13, 247–259. doi: 10.1002/(sici)1099-0984(199907/08)13:4<247::aid-per329<3.0.co;2-7

Furnham, A., and Gasson, L. (1998). Sex difference in parental estimates of their children's intelligence. *Sex Roles* 38, 151–162. doi: 10.1023/A:1018810723493

Furnham, A., Hosoe, T., and Tang, T. (2001). Male hubris and female humility? A cross-cultural study of ratings of self, parental, and sibling multiple intelligence in America, Britain, and Japan. *Intelligence* 30, 101–115. doi: 10.1016/S0160-2896(01)00080-0

Furnham, A., and Rawles, R. (1995). Sex differences in the estimation of intelligence. *J. Soc. Behav. Pers.* 10, 741–748.

Furnham, A., Reeves, E., and Budhani, S. (2002a). Parents think their sons are brighter than their daughters: sex differences in parental self-estimations and estimations of their children's multiple intelligences. *J. Genet. Psychol.* 163, 24–39. doi: 10.1080/00221320209597966

Furnham, A., Shahidi, S., and Baluch, B. (2002b). Sex and culture differences in perceptions of estimated multiple intelligence for self and family. *J. Cross Cult. Psychol.* 33, 270–285. doi: 10.1177/0022022102033003004

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. London: Fontana Press.

Gardner, H. (1999). *Intelligence Reframed: Multiple Intelligences for the 21st Century*. New York, NY: Basic Books.

Gentile, B., Grabe, S., Dolan-Pascoe, B., Twenge, J. M., Wells, B. E., and Maitino, A. (2009). Gender differences in domain-specific self-esteem: a meta-analysis. *Rev. Gen. Psychol.* 13, 34–45. doi: 10.1037/a0013689

Greven, C. U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T., and Plomin, R. (2009). More than just IQ: school achievement is predicted by self-perceived abilities but for genetic rather than environmental reasons. *Psychol. Sci.* 20, 753–762.

Gunderson, E., Ramirez, G., Levine, S. C., and Beilock, S. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles* 66, 153–166. doi: 10.1007/s11199-011-9996-2

Halpern, D. F. (2011). *Sex Differences in Cognitive Abilities*, 4th Edn. Mahwah, NJ: Erlbaum.

Halpern, D. F., Beninger, A. S., and Straight, C. A. (2011). "Sex differences in intelligence," in *The Cambridge Handbook of Intelligence*, eds R. J. Sternberg and S. B. Kaufman (New York, NY: Cambridge University Press), 253–272.

Hines, M. (2011). Gender development and the human brain. *Annu. Rev. Neurosci.* 34, 69–88. doi: 10.1146/annurev-neuro-061010-113654

Hines, M. (2015). Early androgen exposure and human gender development. *Biol. Sex Differ.* 6:3. doi: 10.1186/s13293-015-0022-1

Hirschy, A. J., and Morris, J. R. (2002). Individual differences in attributional style: the relational influence of self-efficacy, self-esteem, and sex role identity. *Pers. Individ. Differ.* 32, 183–196. doi: 10.1016/S0191-8869(01)00003-4

Hogan, H. W. (1978). IQ self-estimates of males and females. *J. Soc. Psychol.* 106, 137–138. doi: 10.1080/00224545.1978.9924160

Huberty, C. J., and Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychol. Bull.* 105, 302–308. doi: 10.1037/0033-2909.105.2.302

Jodl, K. M., Michael, A., Malanchuk, O., Eccles, J. S., and Sameroff, A. (2001). Parents' roles in shaping early adolescents' occupational aspirations. *Child Dev.* 72, 1247–1266. doi: 10.1111/1467-8624.00345

Johnson, D. W., Johnson, R., and Anderson, D. (1983). Social interdependence and classroom climate [Article]. *J. Psychol.* 114, 135–142. doi: 10.1080/00223980.1983.9915406

Jussim, L., and Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: knowns and unknowns, resolved and unresolved controversies. *Pers. Soc. Psychol. Rev.* 9, 131–155.

Kagan, J. (1964). "Acquisition and significance of sex-typing and sex-role identity," in *Review of Child Development Research*, Vol. 1, eds M. L. Hoffman and L. W. Hoffman (New York, NY: Russell Sage Foundation), 137–168.

Kajonius, P. J. (2014). Honesty–humility in contemporary students: manipulations of self-image by inflated IQ estimations. *Psychol. Rep.* 115, 311–325.

Kohlberg, L., and Ullian, D. (1974). "Stages in the development of psycho-sexual concepts and attitudes," in *Sex Differences in Behavior*, eds R. C. Friedman, R. M. Richart, and R. Vande Wiele (New York, NY: Wiley), 209–222.

Kollmayer, M., Schober, B., and Spiel, C. (2018). Gender stereotypes in education: development, consequences, and interventions. *Eur. J. Dev. Psychol.* 15, 361–377.

Kornilova, T. V. (2009). Academic achievement in college: the predictive value of subjective evaluations of intelligence and academic self-concept. *Psychol. Russ. State Art* 2, 309–326.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *J. Pers. Soc. Psychol.* 77, 221–232. doi: 10.1037/0022-3514.77.2.221

Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* 77, 1121–1134. doi: 10.1037//0022-3514.77.6.1121

Lytton, H., and Romney, D. M. (1991). Parents' differential socialization of boys and girls: a meta-analysis. *Psychol. Bull.* 109, 267–296. doi: 10.1037/0033-2909.109.2.267

Maccoby, E. E. (1990). The role of gender identity and gender constancy in sex-differentiated development. *New Dir. Child Adolesc. Dev.* 1990, 5–20. doi: 10.1002/cd.23219904703

Martin, C. L., and Ruble, D. N. (2004). Children's search for gender cues: cognitive perspectives on gender development. *Curr. Dir. Psychol. Sci.* 13, 67–70. doi: 10.1111/j.0963-7214.2004.00276.x

McCrae, R. R., and Costa, P. T. (1985). Updating Norman's "adequacy taxonomy": intelligence and personality dimensions in natural language and in questionnaires. *J. Pers. Soc. Psychol.* 49, 710–721.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., and Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: a meta-analysis. *Pers. Psychol.* 60, 63–91. doi: 10.1111/j.1744-6570.2007.00065.x

McGeown, S. P., Goodwin, H., Henderson, N., and Wright, P. (2011). Gender differences in reading motivation: does sex or gender identity provide a better account? *J. Res. Read.* 35, 328–336. doi: 10.1111/j.1467-9817.2010.01481.x

Mirjalili, R. S., Farahani, H. A., and Akbari, Z. (2011). Self-esteem as moderator of the relationship between self-estimated general intelligence and psychometric intelligence. *Procedia Soc. Behav. Sci.* 30, 649–653. doi: 10.1016/j.sbspro.2011.10.126

Most, D. E. (2008). Patterns of doctoral student degree completion: a longitudinal analysis. *J. Coll. Stud. Ret.* 10, 171–190.

Nash, S. C. (1979). "Sex role as a mediator of intellectual functioning," in *Sex-Related Differences in Cognitive Functioning: Developmental Issues*, eds M. A. Wittig and A. C. Petersen (New York, NY: Academic Press), 263–302.

Neisser, U. C. (1979). The concept of intelligence. *Intelligence* 3, 217–227. doi: 10.1016/0160-2896(79)90018-7

Neisser, U. C., Boodoo, G., Bouchard, T. J. Jr., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: knowns and unknowns. *Am. Psychol.* 51, 77–101. doi: 10.1037/0003-066X.51.2.77

Nosek, B. A., Banaji, M. R., and Greenwald, A. G. (2002). Math= male, me= female, therefore math≠ me. *J. Pers. Soc. Psychol.* 83, 44–59. doi: 10.1037//0022-3514.83.1.44

Pajares, F., and Valiante, G. (2001). Gender differences in writing motivation and achievement of middle school students: a function of gender orientation? *Contemp. Educ. Psychol.* 26, 366–381.

Paulhus, D. L., Lysy, D. C., and Yik, M. S. M. (1998). Self-report measures of intelligence: are they useful as proxy IQ tests? *J. Pers.* 66, 525–554. doi: 10.1111/1467-6494.00023

Phillipson, S., and Phillipson, S. N. (2007). Academic expectations, belief of ability, and involvement by parents as predictors of child achievement: a cross-cultural comparison. *Educ. Psychol.* 27, 329–348.

Pinquart, M., and Ebeling, M. (2019). Parental educational expectations and academic achievement in children and adolescents—a meta-analysis. *Educ. Psychol. Rev.* 32, 463–480. doi: 10.1007/s10648-019-09506-z

Preacher, K. J., and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav. Res. Methods Instrum. Comput.* 36, 717–731. doi: 10.3758/BF03206553

Rammstedt, B., and Rammsayer, T. H. (2002). Gender differences in self estimated intelligence and their relation to gender role orientation. *Eur. J. Pers.* 16, 369–382. doi: 10.1002/per.454

Reilly, D. (2019). Gender can be a continuous variable, not just a categorical one: comment on Hyde, Bigler, Joel, Tate, and van Anders (2019). *Am. Psychol.* 74, 840–841. doi: 10.1037/amp0000505

Reilly, D., and Hurem, A. (in press). "Designing equitable STEM education: guidelines for parents, educators, and policy-makers to reduce gender/racial achievement gaps," in *International Encyclopedia of Education Volume 13: STEM and Beyond*, 4th Edn. Vol. 13, eds X. Liu and L. Wang (Amsterdam: Elsevier).

Reilly, D., and Neumann, D. L. (2013). Gender-role differences in spatial ability: a meta-analytic review. *Sex Roles* 68, 521–535. doi: 10.1007/s11199-013-0269-0

Reilly, D., Neumann, D. L., and Andrews, G. (2016). Sex and sex-role differences in specific cognitive abilities. *Intelligence* 54, 147–158. doi: 10.1016/j.intell.2015.12.004

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Rosenthal, R. (1995). Critiquing Pygmalion: a 25-year perspective. *Curr. Dir. Psychol. Sci.* 4, 171–172.

Rosenthal, R., and Jacobson, L. (1968). Pygmalion in the classroom. *Urban Rev.* 3, 16–20. doi: 10.1007/bf02322211

Sinclair, S. J., Blais, M. A., Gansler, D. A., Sandberg, E., Bistis, K., and LoCicero, A. (2010). Psychometric properties of the Rosenberg Self-Esteem Scale: overall and across demographic groups living within the United States. *Eval. Health Prof.* 33, 56–80. doi: 10.1177/0163278709356187

Snow, R. E. (1995). Pygmalion and intelligence? *Curr. Dir. Psychol. Sci.* 4, 169–171.

Solbes-Canales, I., Valverde-Montesino, S., and Herranz-Hernández, P. (2020). Socialization of gender stereotypes related to attributes and professions among young Spanish school-aged children. *Front. Psychol.* 11:609. doi: 10.3389/fpsyg.2020.00609

Spence, J. T. (1984). "Gender identity and its implications for masculinity and femininity," in *Nebraska Symposium on Motivation: Psychology and Gender*, ed. T. B. Sonderegger (Lincoln, NE: University of Nebraska Press), 59–95.

Spence, J. T., Helmreich, R. L., and Stapp, J. (1974). The Personal Attributes Questionnaire: a measure of sex role stereotypes and masculinity-femininity. *JSAS Cat. Sel. Doc. Psychol.* 4:127. doi: 10.1037/t02466-000

Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J. Pers. Soc. Psychol.* 69, 797–811. doi: 10.1037/0022-3514.69.5.797

Sternberg, R. J. (2000). "The concept of intelligence," in *Handbook of Intelligence*, ed. R. J. Sternberg (Cambridge: Cambridge University Press), 3–15.

Sternberg, R. J., Conway, B. E., Ketron, J. L., and Bernstein, M. (1981). People's conceptions of intelligence. *J. Pers. Soc. Psychol.* 41, 37–55. doi: 10.1037/0022-3514.41.1.37

Storage, D., Charlesworth, T. E., Banaji, M. R., and Cimpian, A. (2020). Adults and children implicitly associate brilliance with men more than women. *J. Exp. Soc. Psychol.* 90:104020. doi: 10.1016/j.jesp.2020.104020

Storek, J., and Furnham, A. (2012). Gender and gender role differences in domain-masculine intelligence and beliefs about intelligence: a study with Mensa UK members. *Pers. Individ. Differ.* 53, 890–895. doi: 10.1016/j.paid.2012.05.039

Storek, J., and Furnham, A. (2013). Gender, g, gender identity concepts, and self-constructs as predictors of the self-estimated IQ. *J. Genet. Psychol.* 174, 664–676. doi: 10.1080/00221325.2013.772501

Storek, J., and Furnham, A. (2014). Gender and task confidence as predictors of the Domain-Masculine Intelligence type (DMIQ). *Pers. Individ. Differ.* 69, 43–49.

Svedholm-Häkkinen, A. M., Ojala, S. J., and Lindeman, M. (2018). Male brain type women and female brain type men: gender atypical cognitive profiles and their correlates. *Pers. Individ. Differ.* 122, 7–12. doi: 10.1016/j.paid.2017.09.041

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: an assessment of the accuracy of gender stereotypes. *J. Pers. Soc. Psychol.* 66, 21–36. doi: 10.1037/0022-3514.66.1.21

Syzmanowicz, A., and Furnham, A. (2011). Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: four meta analyses. *Learn. Individ. Differ.* 21, 493–504. doi: 10.1016/j.lindif.2011.07.001

Szymanowicz, A., and Furnham, A. (2013). Gender and gender role differences in self- and other-estimates of multiple intelligences. *J. Soc. Psychol.* 153, 399–423. doi: 10.1080/00224545.2012.754397

Tenenbaum, H. R., and Leaper, C. (2002). Are parents' gender schemas related to their children's gender-related cognitions? A meta-analysis. *Dev. Psychol.* 38, 615–630. doi: 10.1037/0012-1649.38.4.615

von Stumm, S. (2014). Intelligence, gender, and assessment method affect the accuracy of self-estimated intelligence. *Br. J. Psychol.* 105, 243–253. doi: 10.1111/bjop.12031

Whitley, B. E. (1983). Sex role orientation and self-esteem: a critical meta-analytic review. *J. Pers. Soc. Psychol.* 44, 765–778. doi: 10.1037/0022-3514.44.4.765

Wood, W., and Eagly, A. H. (2015). Two traditions of research on gender identity. *Sex Roles* 73, 461–473. doi: 10.1007/s11199-015-0480-2

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., et al. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature* 573, 364–369. doi: 10.1038/s41586-019-1466-y

Yuen, M., and Furnham, A. (2006). Sex differences in self-estimation of multiple intelligences among Hong Kong Chinese adolescents. *High Ability Stud.* 16, 187–199. doi: 10.1080/13598130600618009

Zhang, Y., and Gong, Y. (2001). Self-estimated intelligence and its related factors. *Chin. J. Clin. Psychol.* 9, 193–195.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Education

ORIGINAL RESEARCH
published: 10 February 2022
doi: 10.3389/feduc.2022.820321

Does Instructional Quality Impact Male and Female University Students Differently? Focusing on Academic Stress, Academic Satisfaction, and Mental Health Impairment

Charlott Rubach[1]*, *Luise von Keyserlingk*[1,2], *Sandra D. Simpkins*[1] and *Jacquelynne S. Eccles*[1]

[1]School of Education, University of California, Irvine, Irvine, CA, United States, [2]Center for Research on Education and School Development, TU Dortmund University, Dortmund, Germany

OPEN ACCESS

Edited by:
Caterina Fiorilli,
Libera Università Maria SS. Assunta University, Italy

Reviewed by:
Diego Boerchi,
Catholic University of the Sacred Heart, Italy
Luciano Romano,
Libera Università Maria SS. Assunta, Italy

***Correspondence:**
Charlott Rubach
crubach@uci.edu

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: 22 November 2021
Accepted: 03 January 2022
Published: 10 February 2022

Citation:
Rubach C, von Keyserlingk L, Simpkins SD and Eccles JS (2022) Does Instructional Quality Impact Male and Female University Students Differently? Focusing on Academic Stress, Academic Satisfaction, and Mental Health Impairment. Front. Educ. 7:820321. doi: 10.3389/feduc.2022.820321

Gender differences in university students' well-being and mental health are prominent concerns in higher education. During the COVID-19 pandemic, male and female students have reported specific stressors that have impacted their well-being and mental health, including difficulty concentrating, concerns about academic performance, and classroom workload. All of these stressors could be mitigated by instructional quality in courses. This study sought to better understand the associations between instructional quality and mental health impairment, i.e., poor mental health and high psychological distress, among male and female undergraduate students during the COVID-19 pandemic. We asked whether perceived instructional quality has a protective effect on students' mental health with regard to academic stress and academic satisfaction across genders. We used longitudinal data from an ethnically diverse sample of 209 students (68% females, 82% freshmen, 50% Asian, 32% Hispanic, 13% White, 5% other) from a public university in Southern California, United States. Data were assessed during the winter and spring quarters of the academic year 2019–2020, i.e., before and after the outbreak of the COVID-19 pandemic in the US. Associations between instructional quality and students' mental health impairment did not differ across genders. The findings indicated that perceived instructional quality at the beginning of the spring quarter 2020 was indirectly related to male and female students' mental health impairment at the end of this quarter. This association was mediated by academic satisfaction. This finding points to a protective effect of instructional quality on students' mental health. However, no effect was found concerning changes to mental health. Gender differences occurred in the link between academic stress and mental health impairment. Academic stress was a stronger predictor of mental health impairment for female students compared to male students. Furthermore, for female students alone, academic stress predicted changes in mental health impairment. We discuss practical implications for higher education. First, our study highlighted that instructional quality in higher education courses might lead to academic satisfaction and thereby help protect university students' mental health. Second, higher education might consider providing additional support for (female) students to improve

their stress management. We argue that improving and enhancing the academic environment are more important than reducing the burden of stressors.

Keywords: gender differences, higher education, instructional quality, mental health impairment, academic stress, academic satisfaction

INTRODUCTION

University students' well-being, mental health, and interrelated factors like stress are prominent concerns in higher education. An alarming percentage of university students report high rates of stress, depressive symptoms, or anxiety as a consequence of multiple stressors (Eisenberg et al., 2013). The unprecedented circumstances of the COVID-19 pandemic have posed additional risks to university students, compelling them to balance an even greater number of stressors simultaneously, consequently impacting their mental health (Son et al., 2020; Smith et al., 2021). Female students especially have reported greater mental health impairment, i.e., poor mental health and higher psychological distress, during the pandemic than male students (Elmer et al., 2020).

During the COVID-19 pandemic, students have experienced several stressors, such as difficulty concentrating, concerns about academic performance, and classroom workload (Son et al., 2020; Smith et al., 2021; Usher et al., 2021). Instructors may be able to help their students cope with these stressors by improving the quality of their instruction. Instructional quality could thus be viewed as a protective factor, one that could mitigate academic stress and mental health impairment. However, to date, little research has been conducted on the extent to which instructional quality alleviates the adverse effects of academic stressors, enhances academic satisfaction, and, consequently, stabilizes or improves the mental health of university students, especially during the COVID-19 pandemic. Furthermore, even though interindividual differences in terms of stressors and protective factors have been indicated (Acharya et al., 2018; Rubach et al., 2020), gender differences in these regards have been examined less frequently. Accordingly, the current study sought to better understand the associations between instructional quality, academic stress, academic satisfaction, and mental health impairment among male and female undergraduate students during the COVID-19 pandemic. We asked whether these students' academic stress and satisfaction mediate the effect of instructional quality on mental health impairment. Furthermore, we aimed to determine whether the instructional quality of online university courses has a protective effect on students' mental health across genders, i.e., if instructional quality is associated with reduced mental health impairment or has positive impacts on the mental health of male and female students.

We believe that our findings are highly valuable for higher education. For instance, our results might help faculty to become better informed about how to maintain and improve the mental health of their male and female students and, in doing so, ensure more equitable academic development for all of their students. Furthermore, our results highlight the need to provide high-quality instruction programs for faculty.

Relation between Instructional Quality, Academic Stress, Academic Satisfaction, and Mental Health Impairment

The Social Environments' Impact on Mental Health: A Theoretical Overview

The main effect model (Cohen et al., 2000) describes the relations between the social environment and individuals' mental health. According to this theory, social interactions and support impact mental health through 1) health-relevant biological influences, 2) health-promoting behaviors, and 3) psychological states, i.e., emotions and cognitions (see Cohen et al., 2000). Individuals' psychological states, such as academic stress and academic satisfaction, mediate the relations between social interactions and mental health. Additionally, these relations can be expected to be bi-directional in that individuals' mental health also impacts the social environment. Based on the tenets of the main effect model (Cohen and Wills, 1985; Cohen et al., 2000), our study focused on the bi-directional associations between instructional quality (indicator of social interaction) and university students' mental health impairment, academic stress, and academic satisfaction (indicator of psychological states, see **Figure 1**).

The Concept of Instructional Quality

Guided by the main effect model (Cohen and Wills, 1985; Cohen et al., 2000), we assumed that university students who attend classes in which the quality of instruction is higher should experience less stress and more satisfaction and, thereby, less severe mental health impairment. The goal, then, was to determine whether this assumption is actually true.

Many scholars have asked about how teaching positively impacts students' academic development (see Devine et al., 2013; Roksa et al., 2016). Such questions have included discussions about instructional quality, which has been defined as multidimensional (Ainley and Carstens, 2018). Two international frameworks (Klieme et al., 2009 (basic dimensions of instructional quality); Pianta and Hamre, 2009 (conceptual framework for classroom interactions)) predict that instructors who use strategies related to classroom management/classroom organization, student (emotional) support, and cognitive activation/instruction provide an effective and cognitively stimulating academic environment. Our study investigated the impact of these three dimensions, which we labeled "classroom management," "student support," and "cognitive activation," on students' academic stress, academic satisfaction, and mental health impairment. The dimension "classroom management" refers to instructional strategies aimed at effectively managing learning and avoiding class disruptions by organizing, monitoring, and managing various

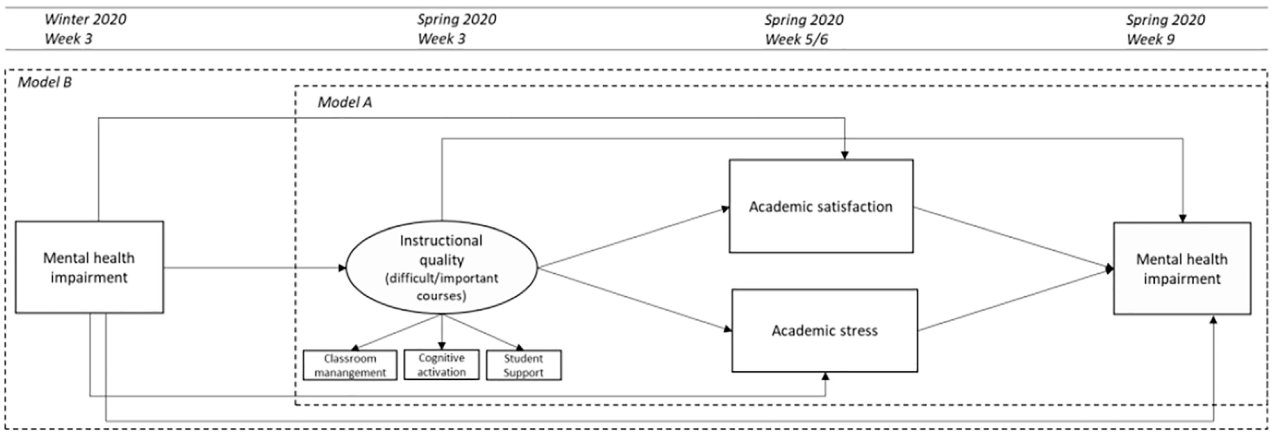


FIGURE 1 | Theoretical tested model.

class settings related to time, resources, assignments, and rules (Praetorius et al., 2018). Classroom management also includes clear instructions, e.g., clearly stated dates, deadlines, and learning goals (Klieme et al., 2009; Pianta and Hamre, 2009). The dimension "student support" and related instructional strategies seek to create a positive learning climate and positive relationships (both teacher-to-student and student-to-student relationships) to improve teachers' and students' motivational beliefs and emotional well-being in class (Praetorius et al., 2018). The dimension "cognitive activation" includes instructional strategies designed to stimulate and support students' cognitive processes and construct and reinforce conceptual understandings and relevant content knowledge for all students (Praetorius et al., 2018). These frameworks have been used in international large-scale assessments—e.g., TALIS 2018, PISA 2012, or TIMSS 2000—to conceptualize instructional quality (see Praetorius et al., 2018). Previous research supports the assumption that all three dimensions are positively related to students' academic development, e.g., students' well-being, motivational beliefs, and performance (Dorfner et al., 2018; Kunter et al., 2008; Rubach and Lazarides, 2021).

The COVID-19 pandemic has prompted rapid changes in the personal and professional lives of both faculty and students at universities. Within a short period of time, courses shifted to an online format, referred to as Emergency Remote Teaching (ERT, Hodges et al., 2020). This unexpected and abrupt transition resulted in numerous changes to how courses were taught. Instructors had to teach and students had to learn in an online environment reliant upon various technologies and digital resources. In this environment, the success of online courses depended on, for example, the digital literacy of instructors and students, instructors' experience administering online courses, and instructors' knowledge of the pedagogical benefits and limitations of online courses (Ferri et al., 2020; Mishra et al., 2020; Adedoyin and Soykan, 2021; Brunetto et al., 2021; Lemay et al., 2021).

Teachers taught their courses either synchronously or asynchronously, or combined both approaches (hybrid format). Most instructors used learning management systems (LMS), video platforms and offered live instruction, delivered teacher-centered presentations, and provided learning materials in the form of short videos or digital texts (Mishra et al., 2020; Lemay et al., 2021). The issue here was that traditional, well-known methods were adapted to an online environment without incorporating the established benefits of online instruction (Adnan and Anwar, 2021).

With respect to the three dimensions of instructional quality, instructors and students reported changes in teaching and learning in online courses compared to in-person courses. These changes were both beneficial and challenging (see Khan, 2021). One challenge noted concerning the shift to remote teaching was that instructors had no or limited experience and knowledge related to classroom management in online courses, such as what rules to implement in video calls, or how to effectively structure courses, or how to monitor student learning and thwart disruptive behaviors (Brunetto et al., 2021). In the online environment, student support was also challenging for both students and faculty. Both groups reported a lack of personal interaction, low student engagement in class, and altered communication processes, such as more asynchronous individual communication and more synchronous group communication (Ferri et al., 2020; Khan, 2021). Another observed challenge was low motivation among students and ineffective methods to individually support students' learning processes (Ferri et al., 2020; Mishra et al., 2020; Lemay et al., 2021). Cognitive activation was also perceived as a challenge: Teachers reported a decline in the quality of student work and felt that they were overwhelming their students (Lemay et al., 2021). Regardless of these challenges, however, the quality of instruction in the online environment nonetheless had an impact on student development (see Aristovnik et al., 2020; Usher et al., 2021; Yu et al., 2021).

How Does Instructional Quality Impact Students' Academic and Personal Outcomes?

Empirical evidence on links between instructional quality as experienced by students and their experience of academic stress, academic satisfaction, and mental health impairment are described in the following. Furthermore, we discuss the indirect effect of instructional quality on students' mental health impairment as mediated by students' academic stress and academic satisfaction.

Instructors who implement classroom management, student support, and cognitive activation strategies in their course can protect their students from becoming frustrated or confused, which can in turn positively impact students' academic satisfaction and reduce their stress (Cassel, 1976; Cobb, 1976; Rubach and Lazarides, 2021). Particularly in the first quarter following the outbreak of the COVID-19 pandemic and implementation of ERT, higher-quality instruction became particularly important as students were being exposed to multiple stressors, some of which were unprecedented, and instructors were in some ways responsible for mitigating these stressors. One university student stated that "the teacher's effectiveness is key in online courses and probably even more so than traditional courses, because online courses can be just a string of homework assignments throughout the whole semester" (Smith et al., 2021, p. 790).

Overall, research supports the conclusion that instructional quality influences students' academic stress: Courses involving higher workload or producing lower grades than anticipated can be particularly stressful (Acharya et al., 2018; Son et al., 2020). During the COVID-19 pandemic, students have reported that lower-quality instruction and slow learning progress have reduced their motivation to learn and increased their tendency to procrastinate (Son et al., 2020). Therefore, we predicted that higher-quality instruction would lowered stress among students during the COVID-19 pandemic. Supporting this prediction, studies have shown that students who have been satisfied with the quality of instruction in courses during the COVID-19 pandemic have also experienced less academic stress (Aristovnik et al., 2020; Usher et al., 2021; Yu et al., 2021).

Furthermore, students have reported greater academic satisfaction and more positive emotions when enrolled in classes with high-quality instruction (Sax et al., 2005; Artino, 2008; Lee, 2010; Ralston-Berg et al., 2015; Habe et al., 2021; Holzer et al., 2021). Students have also reported that higher instructional quality is one of the most important features of courses, increasing their satisfaction and well-being (Smith et al., 2021). Further, higher instructional quality likely culminates in more positive emotions and reduces the likelihood of negative emotions in courses (Rubach and Lazarides, 2021).

Guided by the main effect model, we further hypothesized an indirect link between instructional quality and students' mental health impairment as mediated by their association with academic stress and academic satisfaction. As described above, instructional quality is associated with academic stress and academic satisfaction, both of which in turn influence mental health impairment (Acharya et al., 2018; Rezaei et al., 2015; Shankar and Park, 2016; Shi, 2021). Therefore, we assumed that academic stress and academic satisfaction mediate the link between instructional quality a3nd mental health impairment.

How Does Mental Health Impacts Students' Academic Outcomes?

According to the main effect model (Cohen and Wills, 1985; Cohen et al., 2000), researchers must account for the bi-directional effects between instructional quality and students' mental health impairment. Through these bi-directional effects, individuals' mental health can be expected to impact their perceptions of their social environment. Findings have shown that students with poorer mental health perceive their environment more negatively than those with better mental health. For example, students with poorer mental health perceive less support from their teachers (Tinklin et al., 2005; Rubach et al., 2020). In addition, poorer mental health also leads to less satisfaction and more stress if associated mental health impairments have not been adequately treated (Lipson and Eisenberg, 2018; von Keyserlingk et al., 2021). These results support the theorized bi-directional links between social support, psychological states, such as satisfaction and stress, and mental health as described in the main effect model (Cohen and Wills, 1985; Cohen et al., 2000).

Investigating such bi-directional effects is essential for research on the impact of the COVID-19 pandemic on students' academic development. In terms of stress, the COVID-19 pandemic has caused increased stress among university students on multiple levels (von Keyserlingk et al., 2021). The future health impacts of the COVID-19 pandemic remain unknown. Furthermore, changes in social interactions and work/academic settings due to lockdowns and the imposition of pandemic-related safety measures have heightened stress, decreased students' academic satisfaction, and adversely affected mental health (see Aristovnik et al., 2020; Schiff et al., 2020). University students have reported increased stress related to their coursework, have admitted to procrastinating more often, and have decried disruptions to their study-life balance since the outbreak of the COVID-19 pandemic in spring 2020 and subsequent lockdowns in the US (von Keyserlingk et al., 2021). Additionally, as predicted by Cohen et al. (2000), students who had poor mental health before the COVID-19 pandemic have suffered from a worsening decline in their academic stress as the pandemic has dragged on (von Keyserlingk et al., 2021).

There are, however, particular deficiencies in the extent of research on the bi-directional interrelation of instructional quality and students' academic development in the context of higher education. To our knowledge, no study has investigated the bi-directional effects between mental health, instructional quality, academic stress, and academic satisfaction across multiple time points within a single academic quarter. We argue that a better understanding of these associations is critical for determining the relevance of instructional quality

to healthy academic development among students in higher education and for identifying strategies to enhance higher education teaching.

Prevention and Intervention: The Importance of Instructional Quality

This study focused on the relation between instructional quality and students' mental health. Tinklin et al. (2005) noted the significance of finding resources in higher education that would positively impact students' mental health. One important question in this regard is the extent to which instructors can protect students from experiencing mental health impairments. Higher-quality instruction may serve this aim through either prevention or intervention. Prevention is geared toward reducing the risk of negative outcomes through, for example, targeted reinforcement of relevant competencies and beliefs. Doll et al. (2014) claimed that "school classrooms can become resilient communities that provide essential support and guidance so that vulnerable children can learn and be successful." If so, then instructors who use teaching strategies to structure classes, outline clear expectations and deadlines, and offer individual support might prevent or mitigate stress and anxiety among their students concerning, for instance, exams, which, in turn, would promote better mental health. In contrast, intervention is understood as an intentional, proactive method "to interfere with and stop or modify a process" (American Psychological Association, 2020). In the context of instructional quality, instructors could offer feedback tailored to individual students who fail an exam and are experiencing high levels of stress and test anxiety.

This study, however, was designed to investigate the extent to which students are less likely to experience stress and mental health impairment and more likely to be satisfied in a high-quality instructional environment when, for example, they are fully aware of coursework deadlines, receive constructive feedback tailored to their individual performance, and experience learning improvements. First, we focused on the indirect effect of instructional quality on mental health impairment through academic stress and academic satisfaction across multiple time points. Second, we incorporated prior mental health impairment to investigate its association with perceptions of instructional quality, academic stress und academic satisfaction. This approach permitted 1) the investigation of the bi-directional effects between instructional quality and mental health impairment, and 2) the determination of whether instructional quality is associated with changes in university students' mental health.

Gender Differences in the Impact of Instructional Quality and Mental Health

Do gender differences exist in the processes discussed so far? Do female university students respond more strongly to the academic stressors associated with the COVID-19 pandemic than their male peers? Existing evidence suggests that they do: Female students have reported higher academic stress, i.e., study-life balance stress (Kecojevic et al., 2020; Moksnes et al., 2010; von Keyserlingk et al., 2021) and greater mental health impairment (Corrigan et al., 2016; Elmer et al., 2020), but also higher academic satisfaction (Jager and Gbadamosi, 2013; Habe et al., 2021), than male students. Determining mean-level differences across genders can help to identify which groups might benefit most from intervention programs. In addition, examining the underlying mechanisms behind such mean-level differences across groups can allow us to better understand whether the same intervention might be equally effective for both genders. This approach could help to create equitable opportunities for both genders in higher education. We thus argue that it is essential to investigate whether mechanisms of academic development differ between male and female students.

Based on prior findings, we predicted that instructional quality matters more to female students' than male students' mental health and academic satisfaction. For example, in one study, social support by teachers decreased depressive symptoms among female students alone (Rubach et al., 2020); and, in another study, female students who felt they were not taken seriously by their instructors reported lower academic satisfaction (Sax et al., 2005). In contrast to male students, female students also deemed instructional quality to be more relevant to them (Heine and Maddox, 2009; Jung, 2012). Lastly, female students perceived stressors related to coursework to be greater than did their male peers, i.e., stress caused by increased class workload or receiving lower grades than anticipated (Acharya et al., 2018). These results suggest that instructional quality is more salient for female students in terms of their academic development in higher education.

Moreover, gender differences emerged in the associations between academic stress and mental health impairment. For example, stress caused by academic performance was negatively associated with mental health for female students, but not for male students (Zuckerman, 1989; Hubbard et al., 2018). On the other hand, male students coped with stress by becoming more proactive in their stress response (Zuckerman, 1989). Therefore, we predicted a stronger association between stress and decline in mental health for females than for males as well as a stronger association between instructional quality and academic stress for females than for males. Although equitable academic development opportunities for male and female university students should be an a priori goal, the finding that more female than male students suffer from mental health impairment and stress conflicts with this goal (Corrigan et al., 2016; Elmer et al., 2020; Rubach et al., 2020). It is therefore essential to determine which factors protect against mental health impairment in female students. One factor is, of course, instructional quality, as female students, as mentioned above, consider instructional quality to be more important for their academic development than do their male peers (Jung, 2012). As such, we predicted that instructional quality would be more strongly associated with female students' mental health than with that of male students. Since we, as noted above, consider academic stress and academic satisfaction to be mediators, we predicted that the strength of this mediation would be stronger for female students than for their male counterparts.

The Present Study

The present study focused on the indirect link between instructional quality and university students' mental health impairment via academic stress and academic satisfaction during the COVID-19 pandemic. We posed three research questions:

(RQ1) To what extent has instructional quality protected students from mental health impairment via academic stress and academic satisfaction during the COVID-19 pandemic?

Hypothesis 1. We predicted that students' experiences of instructional quality would be indirectly associated with mental health impairment in that perceptions of high-quality instruction in university courses positively affects academic satisfaction and negatively affects academic stress, both of which are associated with lower mental health impairment among students.

(RQ2) To what extent is mental health impairment among students prior to the COVID-19 pandemic associated with their experiences of instructional quality, academic stress, and academic satisfaction during the COVID-19 pandemic?

Hypothesis 2. We predicted that students who had greater mental health impairment prior to the COVID-19 pandemic would report lower-quality instruction and academic satisfaction but higher academic stress during the COVID-19 pandemic.

(RQ3) Do these associations differ across male and female students?

Hypothesis 3. We predicted that instructional quality would be especially important for female students with regard to their academic stress, academic satisfaction, and mental health impairment compared to male students.

In this study, we focused on two types of courses during the spring quarter of 2020, as rated by students: 1) the most difficult course, and 2) the most important course. The students were also asked to explain the rationale for their ratings (see **Supplementary Material**, **Supplementary Table S1**). Concerning the most difficult course, the most common reason given by students for rating the course this way was that its content and tasks were overwhelming. Other reasons concerned the course teaching strategies and methods, its exam policy, low competence beliefs, and low motivation. Concerning the most important course, the most common reason given by students for rating the course this way was that it was a requirement for their major. Other reasons were that the course held personal value or was important for their future career path.

The students were then asked to report on the instructional quality of the most difficult and most important courses. The association between students' experiences of instructional quality in their most difficult and most important courses and their academic stress, academic satisfaction, and mental health impairment was subsequently investigated. This approach allowed us to examine interindividual differences between different courses rather than using only one course for generalization.

MATERIALS AND METHODS

Sample

The data for this study were derived from an the ongoing *Next generation undergraduate success measurement project* (Arum et al., 2021) project with a longitudinal and multi-cohort design at a public university in Southern California, United States. More specifically, we used data from the subproject "Improve Teaching, Motivational Beliefs, and Well-Being in Higher Education" (Rubach, Eccles, Simpkins and Arum, 2019-2021; see https://www.researchgate.net/project/IMPROVE-Teaching-Motivational-Beliefs-and-Well-Being-in-Higher-Education [02.02.2022]). This subproject investigates on the impact of instructional quality on students' positive development in higher education. The study was designed to investigate undergraduates' experiences and successes. It was approved by the university's Institutional Review Board (IRB). Each cohort was followed over the course of two academic year, with students participating in five surveys per year. All undergraduates in their freshman and junior years at the University of California, Irvine were invited to participate in the study via email. Those students who consented to participate in the study were asked to participate in additional, multiple weekly surveys for which they received course credits. Students who agreed to participate in this part of the study completed short weekly surveys throughout the entire academic year, i.e., data were collected for these students on their weekly academic development across the fall, winter, and spring academic quarters. The weekly surveys focused on course-specific and general questions—questions concerning, for example, a diverse range of course-related experiences, such as instructional quality, motivational beliefs, learning behavior, general well-being, and general college experiences (e.g., mental health, social belonging).

In the present study, we used the data from the first cohort, comprising 1,249 students. These data were collected in the 2019-2020 academic year, beginning in September. Data from this cohort was employed to investigate the impact of instructional quality on students' academic stress, academic satisfaction, and mental health in the first quarter after the COVID-19 pandemic lockdown, spring 2020, when university courses had been converted to a remote, virtual format. From the full cohort, a subsample of 353 students participated in weekly surveys across the academic year (fall 2019 to spring 2020).

We focused on data generated by 209 undergraduates in winter and spring 2020 (age: $M = 19.57$, $SD = 5.43$). These students were selected as they had provided complete responses to at least 70% of the items used for this study. Of the subsample, 82% were students in their freshman year (18% were juniors). Additionally, the subsample had the following characteristics: 68% were female, 54.5% were first-generation college students, and 43.1% had a low family income background. Moreover, students were ethnically diverse (50% Asian; 32% Hispanic; 13% White; 5% other). Lastly, the students were enrolled in different majors (e.g., 39% social sciences; 29% life science; 18% STEM fields).

Instruments

The items, factor loadings, and internal reliability for the constructs are listed in the **Supplementary Material**, **Supplementary Table S2**.

Instructional Quality

Undergraduates reported on instructional quality in their most difficult and most important courses in the third week of the spring quarter 2020. In this study, we adapted existing items on instructional quality from PISA 2012 (OECD, 2013) and developed new items based on the three basic dimensions of the instructional quality framework (Klieme et al., 2009; Praetorius et al., 2018). We did not use existing instruments as they were not developed for higher education, and instruments on instructional quality had to be adapted to the particular context, i.e., the school system (see Praetorius et al., 2018). Therefore, we used either adapted or newly developed items to capture the dimensions of instructional quality and multiple subcategories of the three basic dimensions extracted from Praetorius et al. (2018): For classroom management, we included items that assessed the subcategories of clear rules and both routines and (effective) time use. For student support, we included items that measured the subcategory of competence support. For cognitive activation, the subcategories of challenging tasks and questions as well as exploring and activating prior knowledge were captured by the deployed items. Each dimension of instructional quality (classroom management (CM), student support (SS), cognitive activation (CA)) was assessed with three items. The response scale ranged from 1 = *not at all* to 7 = *very much*. High values of instructional quality indicated that students perceived teaching through multiple instructor behaviors in their courses to be of high quality.

Focusing on the instrument's validity, the results of confirmatory factor analyses indicated that the theoretically described three-factor structure (with a higher-order factor of instructional quality) fit the data better than the one-factor structure, in which all items loaded on one factor (important course: higher-order three-factor model: χ^2 24) = 28.84, p = 0.23, AIC = 5467.95, BIC = 5567.79, CFI =.994, RMSEA = 0.03, SRMR = 0.03; one-factor model: χ^2 27) = 62.38, p = 0.00, AIC = 5524.46, BIC = 5614.31, CFI = 0.959, RMSEA = 0.08, SRMR = 0.04 difficult course: higher-order three-factor model: χ^2 24) = 25.54, p = 0.38, AIC = 5745.57, BIC = 5845.55, CFI = 0.999, RMSEA = 0.02, SRMR = 0.02; one-factor model: χ^2 27) = 85.59, p = 0.00, AIC = 5838.90, BIC = 5928.89, CFI = 0.944, RMSEA = 0.10, SRMR = 0.04). Furthermore, the correlations of the three dimensions of instructional quality (CM, SS, CA) with students' most difficult and most important courses (CM—SS: 0.79*–0.84*; CM—CA: 0.78*–0.84*; CA—SS: 0.80*–0.82*) were similar to the coefficients reported in previous studies (CM—SS: 0.35*–0.69*; CM—CA: 0.49*–0.70*; CA—SS: 0.49*–0.66*; see Kunter et al., 2008; Holzberger et al., 2013). The instrument demonstrated predictive validity as the scales were associated with students' academic outcomes, i.e., their competence beliefs (CB) and subjective task value (STV) in their most important course and most difficult course (CB: |0.24*–0.39*|, STV: |0.46*–0.50*|); these correlations evidenced the same effect range as that reported in previous studies (CB: |0.26*–42*|, STV: |0.32*–0.55*|; Sánchez-Rosas and Esquivel, 2016; Ruiz-Alfonso et al., 2021).

In our analyses, instructional quality was included as a higher-order factor regressed on the manifest scales of classroom management, student support, and cognitive activation.

Academic Stress

Academic stress was operationalized with three items adapted from the University Stress Scale by Stallman and Hurst (2016). Students were asked in the fifth week of the spring quarter 2020 how often in the past 7 days they had experienced stress because of 1) academic/coursework demands, 2) procrastination, and 3) study-life balance. The response scale ranged from 0 = *never* to 7 = *every day* and assessed the frequency of students' perceived academic stress.

Academic Satisfaction

Three items assessed students' satisfaction within their academic environment in week 6 of the spring quarter 2020. These items were developed for this study. Students were asked how satisfied they were with their courses, with their courses' intellectual quality, and with the amount of support they received for learning in their courses. A slider from 0 = *not at all* to 100 = *very much* was used for these items. High values indicated that the students were very much satisfied with, for example, the learning support provided in all courses in which they were enrolled.

Mental Health Impairment

Students' mental health impairment was operationalized with the K10 screening instrument for non-specific psychological distress by Kessler et al. (2002). This established instrument is an indicator for screening mood and anxiety disorders (Furukawa et al., 2003). We used 10 items to ask students how often they had experienced symptoms of psychological distress, such as feeling nervous, hopeless, depressed, or restless, in the third week of the winter quarter 2020 and in the ninth week of the spring quarter 2020. The response scale ranged from 0 = *none of the time* to 4 = *all of the time*. The instrument used in the winter quarter 2020 assessed psychological distress in a range of 7 days, whereas the instrument used in the spring 2020 quarter assessed psychological distress in a range of 30 days. These items were transformed into a sum score with higher values indicating higher psychological distress during the last 7 or 30 days.

Statistics

For all analyses, we used SPSS version 26 as well as MPlus version 8 (Muthén and Muthén, 2016). This study investigated differential associations between instructional quality, academic stress, academic satisfaction, and undergraduates' mental health impairment across female and male students. Guided by IBM SPSS Statistics (2020), differences in scale ranges for instructional quality, academic stress, and academic satisfaction were transformed with linear interpolation into a scale ranging from 1 to 7.

Sample sizes were small, with fewer than 100 cases for the group of males. Therefore, for all analyses, path models were estimated. The constructs of academic stress, academic

satisfaction, and mental health impairment were included as manifest indicators. Instructional quality, however, was added as a higher-order factor with the three dimensions as manifest constructs.

As a first step, measurement invariance across gender groups was investigated. We used the approach proposed by Marsh and others (2015); all constructs needed to be strong invariant with equal factor loadings and item intercepts of constructs across genders. We used cut-off criteria for samples smaller than 300 cases as defined by Chen (2007). A change of 0.005 in the comparative fit index (CFI), supplemented by a change of 0.010 in the root mean square error of approximation (RMSEA), can be interpreted as an indicator of invariance across genders. Marsh et al. (2015), however, emphasized that these cut-off values are rough guidelines. The results highlighted partial strong factorial invariance across genders for academic stress and academic satisfaction. Other investigated constructs were strong factorial invariant across genders. The results on measurement invariance are listed in the **Supplementary Material**, **Supplementary Table S3**.

In the following, we describe our stepwise approach of the data analyses guided by our research questions. Related to the first research question, two models (Models A.1 and A.2) investigated the links between instructional quality (week 3), academic stress (week 5), academic satisfaction (week 6), and mental health impairment (week 9, see **Figure 1**). In Model A.1, we included instructional quality in students' most difficult course; whereas in Model A.2, we used instructional quality in students' most important course (week 3). Related to the second research question (Models B.1 and B.2), we investigated whether undergraduates' mental health impairment prior to the COVID-19 pandemic (week 3, winter quarter 2020) was associated with instructional quality (week 3, spring quarter 2020), academic stress (week 5, spring quarter 2020), academic satisfaction (week 6, spring quarter 2020), and mental health impairment (week 9, spring quarter 2020, see **Figure 1**). In Model B.1, we included instructional quality in students' most difficult course; whereas in Model B.2, we included instructional quality in students' most important course (week 3). In all models, the nested data within the most difficult/important course were taken into account (*type = complex*).

We used the multigroup approach and tested whether associations differed across female and male students across all four models regarding the third research question. Statistical differential effects were detected with the Wald χ^2-test in Mplus (Kodde and Palm, 1986). A non-significant test indicates no meaningful differences across genders. Indirect effects were tested in the full model and the multigroup model for female and male students. The fit of the models to our data was evaluated using cut-offs of model fit indicators guided by Klin (2010) and Brown (2015): CFI $\geq$ 0.90 and RMSEA $\leq$ 0.08 for an acceptable model fit, and CFI $\geq$ 0.95, TLI $\geq$ 0.95, and RMSEA $\leq$ 0.06 for a good model fit. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to compare the baseline models with the multigroup models. We included students in this study with data on at least 70% of the items used for these analyses. Missing data were addressed using full-information maximum likelihood (FIML) estimation.

RESULTS

Descriptive Statistics

Intercorrelations of constructs are reported in **Table 1**. Instructional quality in both students' most difficult and important courses (week 3) were positively associated with students' academic satisfaction (week 6) during spring 2020 but were unrelated to students' mental health impairment (week 3, winter 2020; week 9, spring 2020). Instructional quality in students' most difficult courses (week 3) was related to students' academic stress (week 5) during spring 2020. Academic satisfaction (week 6) and academic stress (week 5) were weakly negatively related to each other during spring 2020. Academic satisfaction (week 6) and academic stress (week 5) during spring 2020 were related to mental health impairment (winter 2020, week 3; spring 2020, week 9).

Instructional Quality on Mental Health Impairment Mediated by Academic Stress and Satisfaction

We first describe direct and indirect effects and gender differences for Model A.1 (instructional quality in students' most difficult course, see **Figure 2**) and afterwards for Model A.2 (instructional quality in students' most important course, see **Figure 3**). Again, these models only used data from spring 2020. The model fit indicators for all models, which are listed in **Table 2**, indicated that each multigroup model fit the data better compared to the baseline model without gender differences (see **Table 3**). The results of the Wald χ^2 test on gender differences are reported in the **Supplementary Material**, **Supplementary Table S4**. The results on indirect effects are reported in the **Supplementary Material**, **Supplementary Table S5**.

Instructional Quality in Students' Most Difficult Course (Model A.1)

The results in Model A.1 (see **Figure 2**) indicated that instructional quality in students' most difficult course predicted students' mental health impairment through their academic satisfaction.

In detail, male and female students' academic satisfaction (week 6) was explained by their reported instructional quality in their most difficult courses (week 3). Students' reported instructional quality in their most difficult courses (week 3) was only marginally associated with their academic stress. Students' academic stress (week 5) and academic satisfaction (week 6) predicted their mental health impairment (week 9). Male and female students' academic satisfaction but not academic stress mediated the link between instructional quality in their most difficult courses (week 3) and their mental health impairment (week 9) (ßind = −0.08, *SE* = 0.03, p = 0.04, 95% CI [−0.13; −0.00]). Gender differences occurred, with academic

TABLE 1 | Descriptive statistics of all scales.

	Quarter	Week	Range	Female					Males				
				n	*Min*	*Max*	*M*	*SD*	*n*	*Min*	*Max*	*M*	*SD*
Instr.qual_diff.w3	Spring 2020	3	1–7	141	1.44	7.00	5.03	1.39	65	1.00	7.00	4.67	1.58
Instr.qual_imp.w3	Spring 2020	3	1–7	140	1.89	7.00	5.72	1.10	66	1.00	7.00	4.83	1.61
Stress.w5	Spring 2020	5	1–7	129	1.00	5.25	3.90	1.17	56	0.00	5.25	3.49	1.41
Satisfaction.w6	Spring 2020	6	1–7	136	0.00	5.94	3.44	1.42	62	0.00	5.94	2.88	1.39
Mental.impair.w9	Spring 2020	9	0–40	123	0.00	40.00	13.28	9.70	56	0.00	40.00	11.32	8.89
Mental.impair.w3	Winter 2020	3	0–40	138	0.00	40.00	12.10	9.65	66	0.00	40.00	11.18	9.42

Note. instr.qual_diff, instructional quality in difficult courses; instr.qual_imp, instructional quality in important courses, stress, academic stress, satisfaction, academic satisfaction, mental.impair, mental health impairment (sum score), w, week.

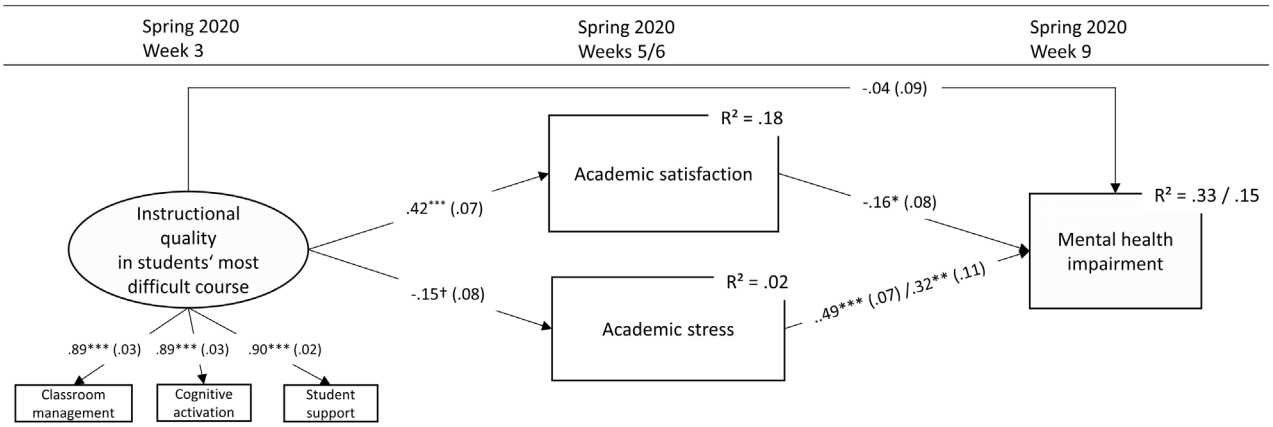


FIGURE 2 | Associations between Instructional Quality in Students' most Difficult Course, Academic Stress, Satisfaction and Mental Health Impairment across Gender (Model A.1). Reported are standardized effects. Gender specific effects for female (n = 137) and male students (n = 62) are reported if the Wald χ^2-test indicated meaningful differences in effects. The order of reported gender effects is: female/males. No gender differences exist if only one standardized effect is reported. *$p \leq 0.10$, **$p \leq 0.05$, ***$p \leq 0.001$.

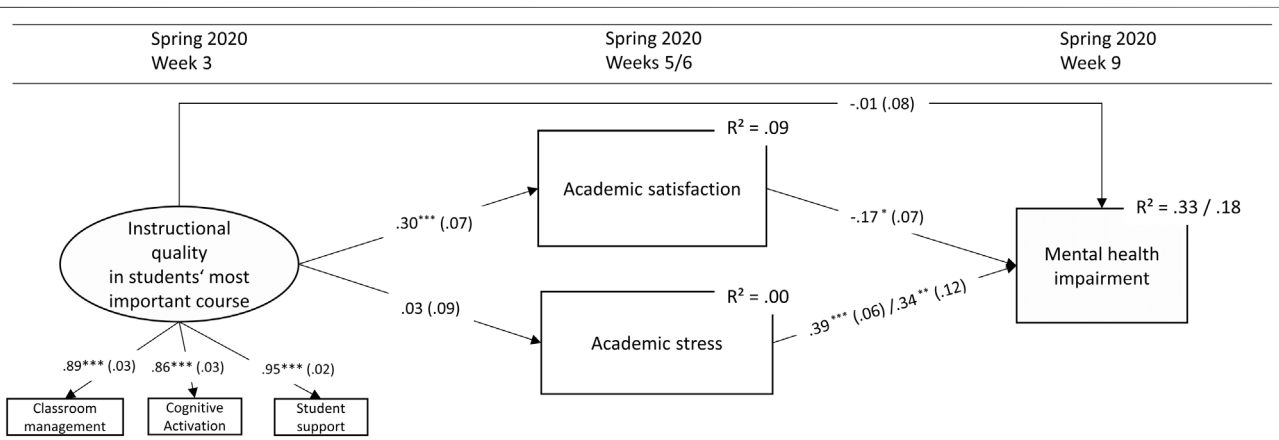


FIGURE 3 | Associations between Instructional Quality in Students' most Important Course, Academic Stress, Satisfaction and Mental Health Impairment across Gender (Model A.2). Reported are standardized effects. Gender specific effects for female (n = 137) and male students (n = 62) are reported if the Wald χ^2-test indicated meaningful differences in effects. The order of reported gender effects is: female/males. No gender differences exist if only one standardized effect is reported. † $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

TABLE 2 | Model fit indices across models.

		χ^2	*df*	*p*	CFI	RMSEA	SRMR	AIC	BIC
Model A.1	Baseline Model	1.655	6	0.95	1.000	0.00	0.01	4389.46	4459.14
	Multigroup Model	16.80	16	0.40	0.998	0.02	0.04	4340.07	4465.78
Model A.2	Baseline Model	2.154	6	0.91	1.000	0.00	0.01	4258.25	4327.93
	Multigroup Model	12.70	16	0.69	1.000	0.00	0.04	4193.69	4319.40
Model B.1	Baseline Model	3.93	8	0.86	1.000	0.00	0.04	4193.58	4276.04
	Multigroup Model	19.46	20	0.49	1.000	0.00	0.04	4166.04	4317.53
Model B.2	Baseline Model	2.14	8	0.98	1.000	0.00	0.01	4068.00	4150.45
	Multigroup Model	13.89	20	0.84	1.000	0.00	0.04	4023.83	4175.33

TABLE 3 | Intercorrelation among analyzed constructs (n = 209).

	Instr.qual_diff.w3	Instr.qual_imp.w3	Stress.w5	Satisfaction.w6	Mental.impair.w9	Mental.impair.w3[b]
Gender[a]	–0.11	–0.31***	–0.14*	–0.17**	–0.10	–0.04
Instr.qual_diff.w3	-	0.37***	–0.15**	0.43***	–0.17**	–0.19**
Instr.qual_imp.w3		-	0.02	0.32***	–0.05	–0.02
Stress.w5				–0.18**	0.35***	0.49***
Satisfaction.w6				-	–0.28***	–0.35***
Mental.impair.w9					-	0.66***

[a]*Note. 0 = female, 1 = male,*

[b]*accessed in Winter 2020, instr.qual_diff, instructional quality in difficult courses; instr.qual_imp, instructional quality in important courses, stress, academic stress, satisfaction, academic satisfaction, ment.imp, mental health impairment (sum score), w, week.* † p < .10, * p < .05, ** p < .01, *** p < .001.

stress (week 5) being a stronger predictor of mental health impairment (week 9) for female as compared to male students.

Instructional quality in students' most difficult courses explained 18% of the variance in their academic satisfaction and 2% of the variance in their academic stress. The variance explanation supported gender differences, with 33% explained variance in mental health impairment for female students, and 15% explained variance in mental health impairment for male students.

Instructional Quality in Students' Most Important Course (Model A.2)

The results in Model A.2 (see **Figure 3**) indicated that instructional quality in students' most important course predicted their mental health impairment through their academic satisfaction.

In detail, higher perceived instructional quality in the most important course (week 3) predicted academic satisfaction (week 6) for both male and female students. In contrast, students' academic stress was not predicted by their reported instructional quality in their most important courses (week 3). Students' academic stress (week 5) and academic satisfaction (week 6) predicted their mental health impairment (week 9). Male and female students' academic satisfaction but not academic stress mediated the link between instructional quality in their most important courses (week 3) and their mental health impairment (week 9) (ßind = –0.05, *SE* = 0.02, p = 0.02, 95% CI [–0.10; –0.01]). Academic stress (week 5) was a stronger predictor of mental health impairment (week 9) for female than for male students.

The explained variance was 9% for students' academic satisfaction and 0% for students' academic stress. The variance explanation supported gender differences, with 33% of the explained variance for mental health impairment for female students and 18% for male students.

Results of Bi-Directional Effects of Instructional Quality and Mental Health Impairment

We first describe direct and indirect effects and gender differences for Model B.1 (instructional quality in students' most difficult course, see **Figure 4**) and afterward for Model B.2 (instructional quality in students' most important course, see **Figure 5**). In these models, we used data from winter 2020 and spring 2020. Model fit indicators indicated that each multigroup model fit the data better compared to the baseline model without gender differences (see **Table 3**). The results of the Wald χ^2 test on gender differences are reported in the **Supplementary Material**, **Supplementary Table S4**. The results on indirect effects are reported in the **Supplementary Material**, **Supplementary Table S6** (Model B.1) and **Supplementary Table S7** (Model B.2).

Instructional Quality in Students' Most Difficult Course (Model B.1)

The results of Model B.1 (see **Figure 4**) indicated that instructional quality in students' most difficult course was not predictive of students' mental health impairment through their academic stress or academic satisfaction when prior mental health impairment was controlled. Hence, instructional quality was not related to changes in students' mental health impairment from winter 2020 to spring 2020.

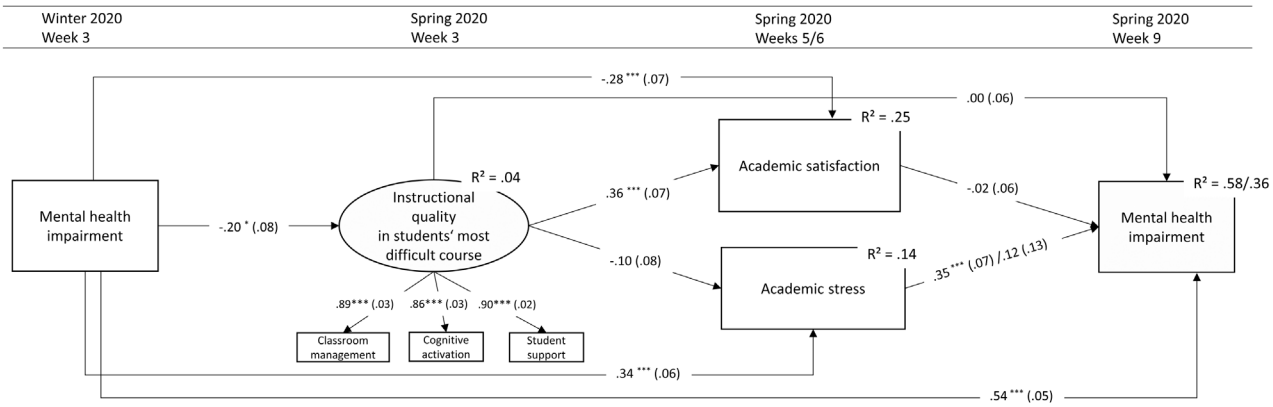


FIGURE 4 | Associations between Instructional Quality in Students' most Difficult Course, Academic Beliefs and Mental Health Impairment Changes Across Gender (Model B.1). Reported are standardized effects. Gender specific effects for female (n = 137) and male students (n = 62) are reported if the Wald χ^2-test indicated meaningful differences in effects. The order of reported gender effects is: female/males. No gender differences exist if only one effect is visualized. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

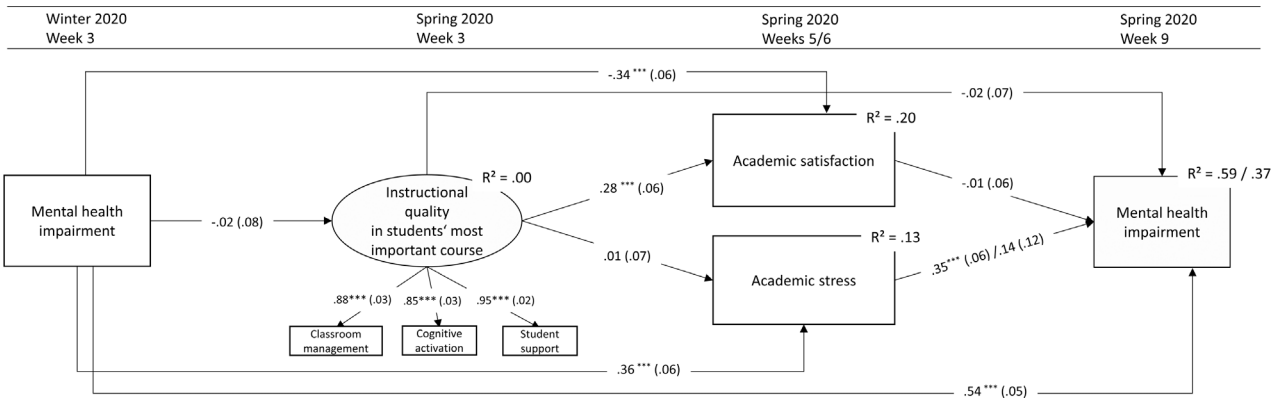


FIGURE 5 | Associations between Instructional Quality in Students' most Important Course, Academic Beliefs and Mental Health Impairment Changes Across Gender (Model B.2). Reported are standardized effects. Gender specific effects for female (n = 137) and male students (n = 62) are reported if the Wald χ^2-test indicated meaningful differences in effects. The order of reported gender effects is: female/males. No gender differences exist if only one effect is visualized. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In detail, the results indicated that for both male and female students, higher mental health impairment (week 3) in winter 2020 predicted lower perceived instructional quality in their most difficult courses in spring 2020 (week 3). Male and female students' higher mental health impairment (week 3) in winter 2020 was also linked to higher academic stress (week 5) and lower academic satisfaction (week 6) in the spring quarter 2020. For all students, the link between mental health impairment in winter 2020 and academic satisfaction in spring 2020 was mediated by their perceived instructional quality in their most difficult course in week 3 of spring 2020 (ßind = −0.07, SE = 0.03, p = 0.04, 95% CI [−0.14; −0.00]). Gender differences occurred with only female students' higher academic stress (week 5), leading to higher mental health impairment from winter 2020 to spring 2020. Indirect effects indicated meaningful gender differences: Female students' academic stress (week 5) mediated the association between their mental health impairment (week 3) in the winter quarter 2020 and their mental health impairment (week 9) in the spring quarter 2020 (ßind = 0.11, SE = 0.03, p = 0.00, 95% CI [0.04; 0.17]).

The explained variance was 4% for students' reported instructional quality in their most difficult course, 14% for academic stress, and 25% for academic satisfaction. The variance explanation supported gender differences, with 58% of the explained variance in mental health impairment for female students and 36% for male students.

Instructional Quality in Students' Most Important Course (Model B.2)

The results in Model B.2 (see **Figure 5**) indicated that instructional quality in students' most important course was not predictive of their mental health impairment through their

academic stress or academic satisfaction when prior mental health impairment was controlled. Instructional quality in students' most important course was not related to changes in their mental health impairment from winter 2020 to spring 2020.

Furthermore, different effects emerged compared to the results for instructional quality in students' most difficult courses. First, male and female students' mental health impairment in the winter quarter 2020 was not associated with their reported instructional quality in their most important courses at the beginning of the spring quarter 2020. However, perceived instructional quality in male and female students' most important courses (week 3) was linked to their academic satisfaction (week 6) but not to their academic stress (week 5) in the spring quarter 2020. Gender differences occurred with female students' higher academic stress (week 5), which led to a positive change in mental health impairment from winter 2020 to spring 2020—higher perceived stress leads to higher mental health impairment. Female students' academic stress (week 5) mediated the association between mental health impairment (week 3) in the winter quarter 2020 and mental health impairment (week 9) in the spring quarter 2020 (ßind = 0.11, $SE = 0.03$, $p = 0.00$, 95% CI [0.06; 0.15]).

The explained variance was 0% for students' reported instructional quality in their most important course, 13% for academic stress, and 20% for academic satisfaction. The variance explanation supported gender differences, with 59% of the explained variance for mental health impairment for female students and 37% for male students.

DISCUSSION

Studies investigating protective factors against mental health impairment in students have often focused on the students themselves. Tinklin and others (2005) argued that the educational environment and resources need to be considered to identify protective factors against students developing mental health impairment in higher education. The present study was focused on the instructional quality of courses as a potential educational resource and protective factor. As students reported higher mental health impairment and academic stress during the COVID-19 pandemic (Elmer et al., 2020; von Keyserlingk et al., 2021), we examined associations between experienced instructional quality, academic stress, academic satisfaction, and mental health impairment across gender groups in the first academic quarter after the beginning of the COVID-19 pandemic in the US. In the following, we discuss our results with regard to our research questions and hypotheses.

Instructional Quality as Protective Factor for Students' Healthy Development

First, we hypothesized that instructional quality would be indirectly associated with students' mental health impairment (see Cassel, 1976; Cohen and Wills, 1985). In summary, the results indicated that academic satisfaction mediated the link between instructional quality and students' mental health impairment during the spring quarter of 2020. However, the experienced instructional quality did not serve as a direct protective factor against mental health impairment during remote teaching in response to the COVID-19 pandemic. This result partially confirms our hypothesis that instructional quality would be linked indirectly to students' mental health during the COVID-19 pandemic (Hypothesis 1). Two different processes define the function of instructional quality for students' healthy academic development: instructional quality can prevent mental health impairment or reduce mental health impairment (intervention). Our results did not demonstrate that instructional quality causes a decrease in mental health impairment - we found no intervening effect.

One explanation for this finding might be that we focused on the first quarter after the beginning of the COVID-19 pandemic lockdown. Empirical studies have reported that students' well-being and mental health decreased with the start of the lockdown (see Son et al., 2020; Smith et al., 2021). We must therefore consider whether instructional quality might have an intervening effect in such challenging times. Furthermore, we focused on instructional quality in ERT. Important factors, such as cognitive activation and student support, were perceived as inadequate or difficult to implement in online settings (see Ferri et al., 2020; Khan, 2021; Lemay et al., 2021). We encourage future studies to replicate our results under different circumstances, e.g., with in-person classes as a reference or in less challenging times.

However, we would argue that courses with overall high instructional quality provide an educational environment that supports the positive psycho-emotional development for university students. Our results showed that higher perceived instructional quality was related to lower mental health impairment at the end of an academic quarter as mediated by higher satisfaction. This result might indicate the preventive function of instructional quality such that students who experience high instructional quality in their courses are more satisfied and are therefore less likely to develop psychological distress. On the other hand, other mediators could be considered to understand the intervening effect of instructional quality on students' mental health impairment, such as self-efficacy (Shankar and Park, 2016).

However, the missing path concerning changes in mental health impairment might indicate that instructional quality cannot serve the same function as specifically designed support and intervention programs. Several studies have reported that especially continuous, formal, and informal social support services help students with mental health impairment (Cohen et al., 2000). As a result, students need professional support and interventions, e.g., mindfulness-based programs or stress management interventions, to learn to regulate their mental health impairment (Bergen-Cico et al., 2013; Bettis et al., 2017).

Another important finding related to the two courses we observed is that instructional quality in courses perceived to be most difficult explained more variance in students' academic satisfaction than instructional quality in courses perceived to be most important. It is well known that college students have experienced multiple stressors during the COVID-

19 pandemic—as such, it might be that high instructional quality in courses perceived to be most difficult is even more important than in courses perceived to be most important, as difficult courses are considered an higher stressor. This might show that instructors have to provide overall higher instructional quality, especially in challenging situations. In this regard, Cassel (1976) emphasized the relevance of improving and enhancing resources rather than reducing the burden of stressors. We strongly encourage higher education and instructors of courses perceived to be difficult to consider to meet students' need for higher instructional quality in these courses.

Unexpectedly, instructional quality was not related to students' academic stress, and thus academic stress did not mediate the association between instructional quality and students' mental health impairment. Even though multiple scholars have highlighted the stressors that can be addressed by instructional quality, i.e., high workload, low motivation and ability to concentrate, instructional quality in one course might not be enough to facilitate a less stressful academic environment for students (Son et al., 2020; Smith et al., 2021; Usher et al., 2021). This study only focused on two courses (the most important and most difficult courses). It might be that instructional quality across all enrolled courses matters with regard to students' academic stress. Indicated stressors, like high workload, low motivation, or difficulty concentrating, were aggregated across all enrolled courses. We must also keep in mind that the beginning of the COVID-19 pandemic, especially the first quarter after it began, was a highly stressful situation for university students. High instructional quality in this period might not be the most important resource needed to decrease academic stress. Therefore, it might be necessary 1) to investigate instructional quality across all courses in which students are enrolled, and 2) to investigate the association between instructional quality and stress with a different sample at a different time.

Furthermore, the missing link between instructional quality, academic stress, and mental health might also suggest that academic stress could moderate the link between instructional quality and mental health impairment. Our study was guided by the main effect model, which proposed that "social resources have a beneficial effect irrespective of whether persons are under stress" (Cohen et al., 2000, p. 11). However, it might be that instructional quality only impacts mental health when students are under a certain degree of stress (see stress-buffering model, Cohen et al., 2000). The extent to which stress functions as a moderator might be of interest to examine in future studies.

Our findings also indicated that higher prior mental health impairment was linked to students' perception of lower instructional quality, but only in the most difficult courses (see RQ 2). Previous studies have found the same results, i.e., the impact of prior mental health impairment on students' perception of their social environment (Tinklin et al., 2005; Rubach et al., 2020). We added to these results, as we investigated difficult and important courses for students. Our results may highlight the difficulty experienced by students with a mental health impairment in perceiving and coping with difficult situations. Such findings may help to improve higher education as they underscore the need to know more about and address the circumstances and conditions of individual students. Students with mental health impairments might need more individual attention to address their negative views of their environment. This argument is in line with the Person-Environment Fit approach in classrooms and the relevance of addressing students' needs in classrooms (Fraser and Fisher, 1983).

Gender Differences in Students' Healthy Development

No gender differences occurred with respect to the importance of instructional quality for students' academic stress, academic satisfaction, or mental health impairment (RQ3). Therefore, Hypothesis 3 was rejected. As described above, instructional quality mattered for male and female students' healthy (academic) development in higher education. We could not replicate the findings that instructional quality is essential for female students' mental health impairment (Rubach et al., 2020). Differences might be explained by the fact that Rubach et al. (2020) studied male and female students' mental health development from 9th to 12th grade. In contrast, our study focused on the beginning of students' higher education careers. Furthermore, we investigated the overall instructional quality, and Rubach and others (2020) focused only on the instructional quality dimension of "student support." This dimension captures social and emotional support from instructors. Future studies might explore the relevance of each dimension of instructional quality to students' healthy academic development in courses.

It is important to note that our sample participated in remote courses. We do not yet know whether instructional quality in in-person classes is equally important for males and females in terms of stress or whether instructional quality in different types of remote courses (e.g., synchronous or asynchronous classes) impacts males and females differently. This might serve as the topic of further investigation.

A gender-specific developmental process revealed in this study was that academic stress was a predictor of increased mental health impairment only in women (see also Zuckerman, 1989). This result indicates gender-specific development in higher education. Questions for further research are as follows: 1) What processes explain this gender-specific association in women? 2) How can this gender-specific association in higher education be addressed? Shankar and Park (2016) discussed whether the association between stress and psychological distress would be moderated by women's self-efficacy or capacity for stress management. Therefore, future research might focus on the gender-specific association between stress and mental health impairment and the relevance of self-efficacy or capacity for stress management. Another question is how to address these results in higher education as related to equitable development opportunities for male and female students. As stress impacts changes in female students' mental health impairment, it might be essential to educate female students in stress management (Bergen-Cico et al., 2013).

Limitation and Future Steps

This study had several limitations. First, due to the sample size, especially the number of males, we used a manifest modeling approach. Manifest conducted models do not control for measurement errors. Therefore, we tested strong invariances across genders for each construct. However, it might be necessary to replicate the findings with a latent structure equation model.

Guided by previous studies, mental health impairment was calculated as a sum score (see Kessler et al., 2002). Fried and Nesse (2015) questioned this approach. They argued that sum scores collapse different symptoms and assume the same weight for each symptom. Fried and Nesse (2015, p. 6) were concerned that a sum score assumes that "two individuals with equal sum-scores may have clinical conditions whose severities differ drastically." Therefore, sum scores discard critical information about individual symptoms and their combination. Future studies might use more differentiated measures of mental health impairment and psychological disorders, primarily when (interindividual) gender differences in mental health impairment are investigated.

In line with established instruments to measure stress (Cohen et al., 1983), students' perceived academic stress was assessed as a frequency score. However, the perceived intensity of stressful events should also be considered. Combining frequency and intensity measures would provide researchers with more fine-grained measures of students' academic stress. Furthermore, as perceived instructional quality was not linked to the frequency of students' academic stress, it might be that instructional quality can prevent students from intensive stressful events. Lastly, this study investigated interindividual differences based on students' gender. It would also be valuable to use an intersectional lens to understand the mental health impairment of university students and its associations with instructional quality, academic stress, and academic satisfaction (see Castillo-Lavergne and Destin, 2019; Rosenfield, 2012). For example, the intersection of ethnicity/race, gender and socioeconomic status has implications for students' mental health (Castillo-Lavergne and Destin, 2019). Uncertainty among working-class Latinx female students predicted their well-being more strongly than in other groups. Multiple groups of marginalized students might benefit from resources such as instructional quality and as such these resources should be considered in future research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article is not publicly available at this point. It is planned to make the data of the entire UCI MUST project publicly available in the future. For further information on this research project and the used data, please contact the research team at uciundergradstudy@uci.edu or visit the website https://education.uci.edu/next-gen-ug-success-project.html.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the IRB UCI. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CR, LV, and JE contributed to conception and design of the study. LV organized the database. CR performed the statistical analysis. CR wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This research was supported by The Andrew W. Mellon Foundation (1806-05902) and the UCI Education Research Initiative.

ACKNOWLEDGMENTS

We thank the whole team of the UCI-MUST study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2022.820321/full#supplementary-material

REFERENCES

Acharya, L., Jin, L., and Collins, W. (2018). College Life Is Stressful Today - Emerging Stressors and Depressive Symptoms in College Students. *J. Am. Coll. Health* 66 (7), 655–664. doi:10.1080/07448481.2018.1451869

Adedoyin, O. B., and Soykan, E. (2020). Covid-19 Pandemic and Online Learning: the Challenges and Opportunities. *Interactive Learn. Environments*, 1–13. doi:10.1080/10494820.2020.1813180

Adnan, M., and Anwar, K. (2020). Online Learning amid the COVID-19 Pandemic: Students' Perspectives. *Online Submission* 2 (1), 45–51. doi:10.33902/jpsp.2020261309

Ainley, J., and Carstens, R. (2018). *Teaching and Learning International Survey (TALIS) 2018 Conceptual Framework*. Paris: OECD Publishing, 187. OECD Education Working Papers. doi:10.1787/799337c2-en

American Psychological Association (2020). *APA Dictionary of Psychology: Intervention*. https://dictionary.apa.org/intervention (Retrieved January 6, 2022).

Aristovnik, A., Keržič, D., Ravšelj, D., Tomaževič, N., and Umek, L. (2020). Impacts of the COVID-19 Pandemic on Life of Higher Education Students: A Global Perspective. *Sustainability* 12 (20), 8438. doi:10.3390/su12208438

Artino, A. R. (2008). Motivational Beliefs and Perceptions of Instructional Quality: Predicting Satisfaction with Online Training*. *J. Comput. Assist. Learn.* 24 (3), 260–270. doi:10.1111/j.1365-2729.2007.00258.x

Arum, R., Eccles, J. S., Heckhausen, J., Orona, G. A., von Keyserlingk, L., Wegemer, C. M., and Yamaguchi-Pedroza, K. (2021). Framework for Measuring Undergraduate Learning and Growth. Change: The Magazine of Higher Learning, 53 (6), 51–59.

Bergen-Cico, D., Possemato, K., and Cheon, S. (2013). Examining the Efficacy of a Brief Mindfulness-Based Stress Reduction (Brief MBSR) Program on

Psychological Health. *J. Am. Coll. Health* 61 (6), 348–360. doi:10.1080/07448481.2013.813853

Bettis, A. H., Coiro, M. J., England, J., Murphy, L. K., Zelkowitz, R. L., Dejardins, L., et al. (2017). Comparison of Two Approaches to Prevention of Mental Health Problems in College Students: Enhancing Coping and Executive Function Skills. *J. Am. Coll. Health* 65 (5), 313–322. doi:10.1080/07448481.2017.1312411

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. Second edition. New York: The Guilford Press. Methodology in the social sciences.

Brunetto, D., Bernardi, G., Andrà, C., and Liljedahl, P. (2021). Teaching as a System: COVID-19 as a Lens into Teacher Change. *Educ. Stud. Math.*, 1–17. doi:10.1007/s10649-021-10107-3

Cassel, J. (1976). The Contribution of the Social Environment to Host Resistance: the Fourth Wade Hampton Frost Lecture. *Am. J. Epidemiol.* 104 (2), 107–123. doi:10.1093/oxfordjournals.aje.a112281

Castillo-Lavergne, C. M., and Destin, M. (2019). How the Intersections of Ethnic and Socioeconomic Identities are Associated With Well-Being During College. *J. Soc. Issues* 75 (4), 1116–1138.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Struct. Equation Model. A Multidisciplinary J.* 14 (3), 464–504. doi:10.1080/10705510701301834

Cobb, S. (1976). Presidential Address-1976. Social Support as a Moderator of Life Stress. *Psychosom Med.* 38 (5), 300–314. doi:10.1097/00006842-197609000-00003

Cohen, S., Kamarck, T., and Mermelstein, R. (1983). A Global Measure of Perceived Stress. *J. Health Soc. Behav.* 24, 385–396.

Cohen, S., and Wills, T. A. (1985). Stress, Social Support, and the Buffering Hypothesis. *Psychol. Bull.* 98 (2), 310–357. doi:10.1037/0033-2909.98.2.310

Cohen, S., Gottlieb, B. H., and Underwood, L. G. (2000). "Social Relationships and Health," in *Social Support Measurement and Intervention: A Guide for Health and Social Scientists*. Editors S. Cohen, L. G. Underwood, and B. H. Gottlieb (New York: Oxford University Press), 3–26. doi:10.1093/med:psych/9780195126709.003.0001

Corrigan, P. W., Kosyluk, K. A., Markowitz, F., Brown, R. L., Conlon, B., Rees, J., et al. (2016). Mental Illness Stigma and Disclosure in College Students. *J. Ment. Health* 25 (3), 224–230. doi:10.3109/09638237.2015.1101056

de Jager, J., and Gbadamosi, G. (2013). Predicting Students' Satisfaction through Service Quality in Higher Education. *Int. J. Manag. Educ.* 11 (3), 107–118. doi:10.1016/j.ijme.2013.09.001

Devine, D., Fahie, D., and McGillicuddy, D. (2013). What Is 'good' Teaching? Teacher Beliefs and Practices about Their Teaching. *Irish Educ. Stud.* 32 (1), 83–108. doi:10.1080/03323315.2013.773228

Doll, B., Brehm, K., and Zucker, S. (2014). *Resilient Classrooms: Creating Healthy Environments for Learning*. Second Edition. New York: Guilford Publications.

Dorfner, T., Förtsch, C., and Neuhaus, B. J. (2018). Effects of Three Basic Dimensions of Instructional Quality on Students' Situational Interest in Sixth-Grade Biology Instruction. *Learn. Instruction* 56, 42–53. doi:10.1016/j.learninstruc.2018.03.001

Eisenberg, D., Hunt, J., and Speer, N. (2013). Mental Health in American Colleges and Universities: Variation across Student Subgroups and across Campuses. *J. Nerv Ment. Dis.* 201 (1), 60–67. doi:10.1097/NMD.0b013e31827ab077

Elmer, T., Mepham, K., and Stadtfeld, C. (2020). Students under Lockdown: Comparisons of Students' Social Networks and Mental Health before and during the COVID-19 Crisis in Switzerland. *PLOS ONE* 15 (7), e0236337. doi:10.1371/journal.pone.0236337

Ferri, F., Grifoni, P., and Guzzo, T. (2020). Online Learning and Emergency Remote Teaching: Opportunities and Challenges in Emergency Situations. *Societies* 10 (4), 86. doi:10.3390/soc10040086

Fraser, B. J., and Fisher, D. L. (1983). Use of Actual and Preferred Classroom Environment Scales in Person-Environment Fit Research. *J. Educ. Psychol.* 75 (2), 303–313. doi:10.1037/0022-0663.75.2.303

Fried, E. I., and Nesse, R. M. (2015). Depression Sum-Scores Don't Add Up: Why Analyzing Specific Depression Symptoms is Essential. *BMC Med.* 13 (1), 1–11.

Furukawa, T. A., Kessler, R. C., Slade, T., and Andrews, G. (2003). The Performance of the K6 and K10 Screening Scales for Psychological Distress in the Australian National Survey of Mental Health and Well-Being. *Psychol. Med.* 33 (2), 357–362. doi:10.1017/s0033291702006700

Habe, K., Biasutti, M., and Kajtna, T. (2021). Wellbeing and Flow in Sports and Music Students during the COVID-19 Pandemic. *Think Skills Creat.* 39, 100798. doi:10.1016/j.tsc.2021.100798

Heine, P., and Maddox, N. (2009). Student Perceptions of the Faculty Course Evaluation Process: An Exploratory Study of Gender and Class Differences. *Res. Higher Educ. J.* 3, 1–10.

Hodges, C., Moore, S., Lockee, B., Trust, T., and Bond, A. (2020). The Difference between Emergency Remote Teaching and Online Learning. *Educause Rev.* 27.

Holzberger, D., Philipp, A., and Kunter, M. (2013). How Teachers' Self-Efficacy is Related to Instructional Quality: A Longitudinal Analysis. *J. Educ. Psychol.* 105 (3), 774.

Holzer, J., Lüftenegger, M., Korlat, S., Pelikan, E., Salmela-Aro, K., Spiel, C., et al. (2021). Higher Education in Times of COVID-19: University Students' Basic Need Satisfaction, Self-Regulated Learning, and Well-Being. *AERA Open* 7, 23328584211003164. doi:10.1177/23328584211003164

Hubbard, K., Reohr, P., Tolcher, L., and Downs, A. (2018). Stress, Mental Health Symptoms, and Help-Seeking in College Students. *PsiChiJournal* 23 (4), 293–305. doi:10.24839/2325-7342.jn23.4.293

IBM SPSS Statistics (2020). Transforming Different Likert Scales to a Common Scale. Available at: https://www.ibm.com/support/pages/transforming-different-likert-scales-common-scale.

John Lemay, D., Doleck, T., and Bazelais, P. (2021). Transition to Online Teaching during the COVID-19 Pandemic. *Interactive Learn. Environments*, 1–12. doi:10.1080/10494820.2021.1871633

Jung, I. (2012). Asian Learners' Perception of Quality in Distance Education and Gender Differences. *Irrodl* 13 (2), 1. doi:10.19173/irrodl.v13i2.1159

Kecojevic, A., Basch, C. H., Sullivan, M., and Davi, N. K. (2020). The Impact of the COVID-19 Epidemic on Mental Health of Undergraduate Students in New Jersey, Cross-Sectional Study. *PLOS ONE* 15, e0239696. doi:10.1371/journal.pone.0239696

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., et al. (2002). Short Screening Scales to Monitor Population Prevalences and Trends in Non-specific Psychological Distress. *Psychol. Med.* 32 (6), 959–976. doi:10.1017/s0033291702006074

Khan, M. A. (2021). -19's Impact on Higher Education: A Rapid Review of Early Reactive Literature. *Educ. Sci.* 11, 421. doi:10.3390/educsci11080421

Klieme, E., Pauli, C., and Reusser, K. (2009). "The Pythagoras Study: Investigating Effects of Teaching and Learning in Swiss and German Mathematics Classrooms," in *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*. Editors T. Janík and T. Seidel (Münster: Waxmann), 137–160.

Kline, R. (2010). *Principles and Practice of Structural Equation Modelling*. 3rd ed. New York: The Guilford Press.

Kodde, D. A., and Palm, F. C. (1986). Wald Criteria for Jointly Testing Equality and Inequality Restrictions. *Econometrica* 54 (5), 1243. doi:10.2307/1912331

Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., and Baumert, J. (2008). Students' and Mathematics Teachers' Perceptions of Teacher Enthusiasm and Instruction. *Learn. instruction* 18 (5), 468–482. doi:10.1016/j.learninstruc.2008.06.008

Lee, J.-W. (2010). Online Support Service Quality, Online Learning Acceptance, and Student Satisfaction. *Internet Higher Educ.* 13 (4), 277–283. doi:10.1016/j.iheduc.2010.08.002

Lipson, S. K., and Eisenberg, D. (2018). Mental Health and Academic Attitudes and Expectations in university Populations: Results from the Healthy Minds Study. *J. Ment. Health* 27 (3), 205–213. doi:10.1080/09638237.2017.1417567

Marsh, H. W., Parker, P. D., and Morin, A. (2015). "Invariance Testing across Samples and Time: Cohort-sequence Analysis of Perceived Body Composition," in *An Introduction to Intermediate and Advanced Statistical Analyses for Sport and Exercise Scientists*. Editors N. Ntoumanis and N. D. Myers (Chicester, West Sussex: Wiley), 101–130.

Mishra, L., Gupta, T., and Shree, A. (2020). Online Teaching-Learning in Higher Education during Lockdown Period of COVID-19 Pandemic. *Int. J. Educ.es. Open* 1, 100012. doi:10.1016/j.ijedro.2020.100012

Moksnes, U. K., Moljord, I. E. O., Espnes, G. A., and Byrne, D. G. (2010). The Association between Stress and Emotional States in Adolescents: The Role of Gender and Self-Esteem. *Personal. Individual Differences* 49 (5), 430–435. doi:10.1016/j.paid.2010.04.012

Muthén, L. K., and Muthén, B. O. (2016). *Mplus 8.1 [Computer Software]*. Los Angeles, CA: Muthén & Muthén.

OECD (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing. doi:10.1787/9789264190511-en

Pianta, R. C., and Hamre, B. K. (2009). Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educ. Res.* 38 (2), 109–119. doi:10.3102/0013189X09332374

Praetorius, A.-K., Klieme, E., Herbert, B., and Pinger, P. (2018). Generic Dimensions of Teaching Quality: the German Framework of Three Basic Dimensions. *ZDM Maths. Educ.* 50 (3), 407–426. doi:10.1007/s11858-018-0918-4

Ralston-Berg, P., Buckenmeyer, J., Barczyk, C., and Hixon, E. (2015). Students' Perceptions of Online Course Quality: How Do They Measure up to the Research. *Il* 4 (1). doi:10.18278/il.4.1.2

Rezaei, T., Bayani, A. A., and Shariatnia, K. (2015). The Prediction of Mental Health Based on Variables of Self-Esteem, Life Satisfaction and hope Among College Students. *Iranian J. Health Educ. Health Promot.* 3 (3), 242–252. http://journal.ihepsa.ir/browse.php?a_id=253andsid=1.andslc_lang=en.

Roksa, J., Arum, R., and Cook, A. (2016). "Defining and Assessing Learning in Higher Education," in *Improving Quality in American Higher Education: Learning Outcomes and Assessments for the 21st Century*. Editors R. Arum, J. Roksa, and A. Cook. 1st ed. (San Francisco: Jossey-Bass), 1–36.

Rosenfield, S. (2012). Triple Jeopardy? Mental Health at the Intersection of Gender, Race, and Class. *Soc. Sci. Med.* 74 (11), 1791–1801.

Rubach, C., Dicke, A.-L., Lazarides, R., Simpkins, S., and Eccles, J. S. (2020). Addressing Adolescents' Depressive Symptoms and Risky Behavior: The Role of Perceived Parents' and Teachers' Social Support. *Jop* 20 (4), 70–101. doi:10.33423/jop.v20i4.3209

C. Rubach and R. Lazarides (2021). "Freude und Angst von Schülerinnen und Schülern. Welche Zusammenhänge existieren zur Individualisierung und Selbstbestimmung im Mathematikunterricht? [Student enjoyment and anxiety. What connections exist to individualization and self-determination in mathematics classrooms]," in *Emotionen in Schule und Unterricht. Bedingungen und Auswirkungen von Emotionen bei Lehrkräften und Lernenden*. Editors C. Rubach and R. Lazarides (Berlin), 169–191.

Ruiz-Alfonso, Z., León, J., Vega-Santana, L., and González, C. (2021). Teaching Quality: Relationships between Students' Motivation, Effort Regulation, Future Interest, and Connection Frequency. *Psicología Educativa* 27 (1), 67–76.

Sánchez-Rosas, J., and Esquivel, S. (2016). Instructional Teaching Quality, Task Value, Self-Efficacy, and Boredom: A Model of Attention in Class. *Revista de Psicología* 25 (2).

Sax, L. J., Bryant, A. N., and Harper, C. E. (2005). The Differential Effects of Student-Faculty Interaction on College Outcomes for Women and Men. *J. Coll. Student Develop.* 46 (6), 642–657. doi:10.1353/csd.2005.0067

Schiff, M., Zasiekina, L., Pat-Horenczyk, R., and Benbenishty, R. (2020). COVID-related Functional Difficulties and Concerns Among university Students during COVID-19 Pandemic: A Binational Perspective. *J. Community Health*, 1–9. doi:10.1080/21683603.2016.1130532

Shankar, N. L., and Park, C. L. (2016). Effects of Stress on Students' Physical and Mental Health and Academic success. *Int. J. Sch. Educ. Psychol.* 4 (1), 5–9. doi:10.1007/s10900-020-00930-9

Shi, B. (2021). Perceived Social Support as a Moderator of Depression and Stress in College Students. *Soc. Behav. Pers* 49 (1), 1–9. doi:10.2224/sbp.9893

Smith, Y., Chen, Y.-J., and Warner-Stidham, A. (2021). Understanding Online Teaching Effectiveness: Nursing Student and Faculty Perspectives. *J. Prof. Nurs.* 37 (5), 785–794. doi:10.1016/j.profnurs.2021.05.009

Son, C., Hegde, S., Smith, A., Wang, X., and Sasangohar, F. (2020). Effects of COVID-19 on College Students' Mental Health in the United States: Interview Survey Study. *J. Med. Internet Res.* 22, e21279. doi:10.2196/21279

Stallman, H. M., and Hurst, C. P. (2016). The University Stress Scale: Measuring Domains and Extent of Stress in University Students. *Aust. Psychol.* 51 (2), 128–134. doi:10.1111/ap.12127

Tinklin, T., Riddell, S., and Wilson, A. (2005). Support for Students with Mental Health Difficulties in Higher Education: the Students' Perspective. *Br. J. Guidance Counselling* 33 (4), 495–512. doi:10.1080/03069880500327496

Usher, E. L., Golding, J. M., Han, J., Griffiths, C. S., McGavran, M. B., Brown, C. S., et al. (2021). Psychology Students' Motivation and Learning in Response to the Shift to Remote Instruction during COVID-19. *Scholarsh. Teach. Learn. Psychol.*. doi:10.1037/stl0000256

von Keyserlingk, L., Yamaguchi-Pedroza, K., Arum, R., and Eccles, J. S. (2021). Stress of university Students before and after Campus Closure in Response to COVID-19. *J. Community Psychol.* 50, 285–301. doi:10.1002/jcop.22561

Yu, L., Huang, L., Tang, H. R., Li, N., Rao, T. T., Hu, D., et al. (2021). Analysis of Factors Influencing the Network Teaching Effect of College Students in a Medical School during the COVID-19 Epidemic. *BMC Med. Educ.* 21 (1), 397–398. doi:10.1186/s12909-021-02825-2

Zuckerman, D. M. (1989). Stress, Self-Esteem, and Mental Health: How Does Gender Make a Difference. *Sex Roles* 20 (7-8), 429–444. doi:10.1007/bf00288001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Education

ORIGINAL RESEARCH
published: 14 February 2022
doi: 10.3389/feduc.2022.819922

The Role of Gender for Teachers' Reactions to Social Exclusion Among Students

Hanna Beißert[1,2], Miriam Staat[3] and Meike Bonefeld[2,4]*

[1] Department of Empirical Educational Research, University of Konstanz, Konstanz, Germany, [2] Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany, [3] Department of Education and Human Development, DIPF Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany, [4] Research Group for Quality of Instruction in Heterogeneous Contexts, School of Social Sciences, University of Mannheim, Mannheim, Germany

Social exclusion, i.e., being kept apart from others and not being allowed to join, is a common phenomenon at school and can have severe consequences for students' healthy development and success at school. This study examined teachers' reactions to social exclusion among students focusing on the role of gender. Specifically, we were interested in potential effects of gender-specific socialization and social expectations linked to gender for teachers' reactions to social exclusion among students. We used hypothetical scenarios in which a student is being excluded from a study group by other students. We focused on the gender of the teacher (as an observer of exclusion) on the one hand and on the gender of the excluded student on the other hand. In the hypothetical scenarios, we varied the gender of the excluded student by using either a typical female or male name. The study included 101 teachers from different school tracks in Germany (M_{age} = 36.93, *SD* = 9.84; 84 females, 17 males). We assessed teachers' evaluations of the exclusion scenario and their anticipated reactions, i.e., how likely they were to intervene in such a situation and what they would specifically do. As expected, the participating teachers showed a general tendency to reject exclusion among students. This tendency was even more pronounced among female teachers compared to male teachers. Interestingly, these gender differences on the attitudinal side did not translate into differences in teachers' behavioral intentions: for the likelihood to intervene, we did not find any differences based on the gender of the teacher. In terms of the gender of the excluded student, things were different: The gender of the excluded student did not affect teachers' evaluations of the exclusion scenario. Yet, the gender of the excluded was relevant for participants' behavioral intentions. Namely, teachers were less likely to intervene in the scenario if a boy was excluded. These findings are in line with considerations related to gender-specific socialization and social expectations linked to gender. Overall, the study demonstrates that gender is an important aspect in the context of social exclusion and further research should explicitly focus on how socialization and gender expectations can explain these findings.

Keywords: social exclusion, teacher reactions, teacher evaluations, gender differences, gender role expectations, socialization

OPEN ACCESS

Edited by:
Daniela Barni,
University of Bergamo, Italy

Reviewed by:
Francesca Giovanna Maria Gastaldi,
University of Turin, Italy
Angela Page,
The University of Newcastle, Australia

****Correspondence:***
Hanna Beißert
beissert@dipf.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *22 November 2021*
Accepted: *18 January 2022*
Published: *14 February 2022*

Citation:
Beißert H, Staat M and Bonefeld M (2022) The Role of Gender for Teachers' Reactions to Social Exclusion Among Students. Front. Educ. 7:819922. doi: 10.3389/feduc.2022.819922

INTRODUCTION

Being socially excluded threatens the possibility of fulfilling one's psychological needs, for instance, social belonging (Williams, 2009). Recurrent experience of social exclusion can have serious consequences for children's health and wellbeing (Gazelle and Druhen, 2009; Sebastian et al., 2010; Fuhrmann et al., 2019; Jiang and Ngai, 2020), their emotional and social development (Gazelle and Druhen, 2009; Murray-Close and Ostrov, 2009), and even their academic achievement (Buhs et al., 2006).

As children and adolescents spend large parts of their lifetime at school, school has great importance as an environment, in that inclusion and exclusion take place. Given the strong impact of social exclusion on health, wellbeing, and achievement, schools should try to promote relatedness and to prevent exclusion among students. For this, teachers play an important role. It has been shown that teachers' behavior in class can have a strong impact on their students' attitudes regarding exclusion. For instance, Mulvey et al. (2021) found that students who perceived better student–teacher relationships as well as students who reported higher support by their teachers, were more likely to judge exclusion to be wrong and to expect that they would defend victims against exclusion. Additionally, teachers establish norms in class that indicate which behaviors are acceptable and which are not, including in terms of social exclusion. With their reactions to social exclusion among students, teachers transmit messages about their own attitudes regarding exclusion and might with that impact their students' attitudes and behavior as well. Thus, it is important to investigate teachers' reactions to social exclusion.

Teachers as Observers of Social Exclusion

According to Riva and Eck (2016, p. ix), social exclusion can be defined as the "experience of being kept apart from others physically or emotionally". This includes situations in which a person is excluded from conversations or activities by one or several other individuals (Wesselmann et al., 2016). As social exclusion among students is a common phenomenon at school, teachers often witness exclusion situations (Killen et al., 2013). Just as people in other contexts generally tend to reject unsubstantiated social exclusion (Wesselmann et al., 2013), this is also the case for teachers in schools. Several studies using hypothetical scenarios demonstrated that teachers as observers of exclusion among students show a general tendency to reject exclusion (Beißert and Bonefeld, 2020; Grütter et al., 2021; Kollerová and Killen, 2021). Witnessing social exclusion typically induces feelings of empathy with the excluded person (Wesselmann et al., 2013). Several studies found evidence that this is also the case for teachers when they witness exclusion among their students (Grütter et al., 2021; Szekely et al., under review). For instance, in a study by Grütter et al. (2021), the most frequently referenced emotions that teachers reported when reasoning about an exclusion scenario were feeling sad and sympathetic for the excluded student. In line with these findings, we assume that based on empathy with the excluded person, combined with the knowledge about the severe consequences associated with social exclusion, teachers show a general tendency to reject exclusion among students.

Besides these general tendencies, it is of interest whether teachers' reactions to social exclusion might be influenced by characteristics of the target of exclusion (i.e., the excluded student) or by characteristics of the teachers as observers of exclusion. In this context, one important characteristic might be gender. More specifically, the gender of the observing teacher on the one hand, and the gender of the excluded student on the other hand. In the current study, we focus on these two aspects when investigating teachers' reactions to hypothetical exclusion scenarios.

Social Exclusion and the Role of the Teacher's Gender

It has already been shown in different contexts that females tend to evaluate exclusion as more reprehensible than males (Killen and Stangor, 2001; Horn, 2003; Malti et al., 2012; Beißert et al., 2019). This also holds for female teachers (Beißert and Bonefeld, 2020; Beißert et al., 2021).

One possible explanation for this could be gender-specific socialization. Namely, the socialization of girls typically has a stronger focus on harmony and the avoidance of interpersonal struggles (Cross and Madson, 1997; Zahn-Waxler, 2000; Hwang and Mattila, 2019). Moreover, in many families, the harmful consequences of aggressive behaviors are much more addressed in the socialization of girls compared to boys, which might lead to more pronounced feelings of empathy in girls (Smetana, 1989). In line with this, females of different ages have been shown to be more empathic than males (e.g., Rueckert and Naybar, 2008; Schulte-Rüther et al., 2008; D'Ambrosio et al., 2009; Van der Graaff et al., 2014).

Thus, female socialization might lead to stronger feelings of empathy on the one hand and a stronger focus on interdependence, belonging, and community on the other hand. This is in accordance with Bakan's (1966) theory of the two basic dimensions "agency" and "communion" that describe how individuals relate to their social world. The main assumption of this theory is that females and males are differentially socialized in terms of the relative emphasis on agency and communion (Bakan, 1966). Agency refers to an individual striving to assert the self, master the environment, experience competence, and achievement. Whereas communion refers to an individual's desire to cooperate and connect closely with others. While females are typically socialized with a stronger focus on communal goals, the socialization of males has a strong focus on agentic goals. As a consequence, females—being communion-oriented individuals—experience stronger fulfillment through relationships, whereas males as agency-oriented individuals experience fulfillment through achievement of their individual goals (Guisinger and Blatt, 1994). Thus, it seems evident that females value relationships more than males. In line with prior research as well as in accordance with Bakan's theory (1966) and considerations related to gender-specific socialization, we assume that female teachers reject social exclusion more strongly than males.

Social Exclusion and the Role of the Excluded Person's Gender

Not only the gender of the observers of exclusion, in our case teachers, might be relevant. Also, the gender of the excluded student might impact teachers' reactions to social exclusion. To date, there is hardly any research on the gender of the excluded person, especially not in educational contexts. To our knowledge, there is only one study that focused on the role of the excluded person's gender for teachers' evaluations of social exclusion in an educational setting. In this study, Kollerová and Killen (2021) found no differences based on the excluded person's gender in teachers' evaluations of the wrongfulness of the exclusion. However, they found differences in teachers' reasoning revealing that the participants used more moral justifications when reasoning about excluded girls compared to excluded boys.

Yet why should the excluded person's gender be relevant for teachers' reactions to social exclusion? One possible explanation are social expectations linked to gender.[1] Generally, with regard to the two genders, quite different expectations are prevalent. These gender expectations typically impact our thoughts and actions in many ways (Neuburger et al., 2015; Retelsdorf et al., 2015; Mello et al., 2019). Thus, gender expectations might also affect our perception of and reactions to exclusion of boys vs. girls.

Socialization also provides a possible explanatory approach here. Having been socialized throughout our lives, we all have learned and internalized systematically differing expectations regarding males and females. Typical expectations that are relevant in this context are those in line with the assumptions of the aforementioned theory by Bakan (1966): females are usually associated with the dimension of communion; males are traditionally associated with characteristics of agency. That is, we expect girls to strongly value interpersonal affiliation and harmony with others (Spence and Helmreich, 1978; Bem, 1981; Eckes, 2010; Tay et al., 2019). Additionally, in line with traditional gender role expectations, girls are typically perceived as more vulnerable than boys and hence might evoke stronger feelings for care (Stuijfzand et al., 2016). Given that we stereotypically perceive girls—compared to boys—as more communal and more vulnerable beings (Bakan, 1966; Gilligan, 1993; Ely et al., 1998; Eckes, 2010), we might expect that exclusion affects girls more strongly than boys. Based on these considerations on social expectations linked to gender, we assume that exclusion might be perceived as more serious for girls than for boys and in consequence, the exclusion of girls should be rejected more strongly compared to the exclusion of boys.

Current Study

The purpose of this study is to extend prior research on teachers' evaluations of and reactions to social exclusion scenarios by analyzing the role of gender. More specifically, we are interested in potential effects of gender-specific socialization and social expectations related to gender. Thus, we focus on the gender of the observing teacher on the one hand and on the gender of the excluded student on the other hand. Focusing on teachers in the role of observers of exclusion among students, we assessed teachers' evaluations of hypothetical exclusion scenarios. Since particularly teachers' behavior can have an impact on their students, it is not only important to analyze teachers' evaluations of exclusion (which reflect an attitudinal aspect), but also their reactions (which capture a behavioral aspect). Given that it is very difficult to realize naturalistic observational studies in the context of social exclusion, especially at schools, we approach the behavioral aspect by assessing behavioral intentions and want to see whether teachers' evaluations of exclusion translate into respective behavioral intentions. Accordingly, we assess teachers' anticipated reactions and interventions. More precisely, we asked them how likely they were to intervene in such a situation and what they would specifically do. Our main interest was to determine whether the gender of an excluded student and the gender of the teacher as an observer of exclusion are relevant factors for teachers' responses to hypothetical exclusion scenarios.

Based on the aforementioned considerations on gender-specific socialization and gender expectations, we want to examine the following hypotheses:

A. We assume teachers to show a general tendency to reject social exclusion among students and to intervene in exclusion situations among students.
B. We hypothesize that female teachers reject social exclusion more strongly and are more likely to intervene compared to male teachers.
C. We expect that the exclusion of girls will be rejected more strongly compared to the exclusion of boys and the likelihood to intervene will be higher when a girl (vs. boy) is excluded.

As an open question, we want to explore if there are any interactions of the excluded student's gender and the gender of the teachers as observers of exclusion. Further, we want to explore if participants' justifications for their decision to intervene in the situation or not as well as their anticipated specific actions differ between female and male teachers' or depending on the excluded students' gender.

MATERIALS AND METHODS

Participants

The study included 101 teachers from different school tracks in Germany (M_{age} = 36.93, SD = 9.84, range: 22–65, 84 females, 17 males). The working experience of the teachers ranged from under 1 to 42 years (M = 8.16, SD = 8.50) with half of the sample being within their first 5 years of service (median = 5.00 years).

[1] In our study, we focus only on binary gender perceptions as we are interested in gender role expectations associated with the female and male gender. However, we acknowledge, that the conception of gender as binary is a narrow conception that not necessarily reflects the full range of possible gender identifications.

Design and Procedure

The study was conducted as an online survey and participants were recruited via different mailing lists and online groups in social media platforms (e.g., Facebook groups). Moreover, flyers advertising the study were distributed in libraries, schools, and public sites of universities. Participation was voluntary and informed consent was obtained from all participants. The study was conducted in accordance with the ethical guidelines of DGPs (German Psychological Society).

Before starting the actual survey, participants were informed of their data protection rights and learned that participation in the study was anonymous and voluntary. They were also informed, that there were no negative consequences if they decided not to participate or to leave the study early without completing it. Prior to the assessment, participants had to confirm that they were willing to participate in the study and understood the information.

Starting the survey, participants provided demographic information, participants were then presented with a hypothetical exclusion scenario. The study took approximately 10 min per person.

Material

In the hypothetical exclusion scenario, one student was excluded from a study group by its classmates. We varied the gender of this excluded protagonist by presenting either a typical male or female name (Lukas vs. Julia) in the scenario. The names used in the scenarios had been pretested in a former study by Bonefeld and Dickhäuser (2018). The exact wording of the scenario was as follows:

> *While packing up after class in 7th grade, you observe some students making an appointment to study together. Lukas/Julia would like to join the learning group. The other students tell him/her that he/she can't join.*

The study was realized as a between-subjects design. The participants were randomly assigned to the experimental conditions (51 were assigned the version with a female protagonist, 50 to the version with the male protagonist).

Measures

As we wanted to assess not only attitudinal but also behavioral aspects, we assessed participants' evaluations of the exclusion situation on the one hand and their likelihood to intervene in such a situation and the specific actions they would undertake on the other hand. We used a seven-point Likert-type scale consisting of three items to assess the evaluations of the exclusion scenario. Specifically, we asked the participants to rate how (1) not okay/okay, (2) unfair/fair, and (3) unjustifiable/justifiable the scenario was. Based on these three items, a score was created indicating a participant's evaluation of the exclusion (Cronbach's

TABLE 1 | Coding system for justifications of likelihood of intervention and frequencies for each category.

Category	Example	*N*
Need for information	*"Because I would like to find out why the group does not want to work with Julia.,"* *"I want to find out more about the situation."*	22
Children's autonomy	*"Extracurricular activities do not concern me.,"* *"It is a private matter of the students."*	12
Empathy for the victim/Avoid psychological harm	*"Because I feel sad for her, it is not nice to be excluded."* *"I want to avoid mobbing."*	8
Social norms of inclusion and cooperation	*"Exclusion is never an option.,"* *"Nobody should be excluded.,"* *"Cooperation and cohesion are important values."*	23
Other	Meaningful, but single statements	8
Undifferentiated	Meaningless statements	3

This question was answered by 73 participants.

TABLE 2 | Coding system for specific actions and frequencies for each category.

Category	Example	*N*
Ask for reasons	*"I would ask the students about the reason for the rejection."* *"I would try to find out why she can't join."*	41
Conversation	*"I would want to talk with them.,"* *"I would talk with them in the group and if necessary, we can have a private conversation."*	29
Inclusion-oriented behavior	*"Appeal to students to include Julia.,"* *"I would point out the behavior that excluding a person is not good."*	16
Find alternative solution for excluded student	*"I would help the student to find another group for studying."* *"For the concerned student, I would make alternative suggestions. Perhaps there is another classmate who also cannot easily find study partners."*	8
Other	Meaningful, but single statements	2
Undifferentiated	Meaningless statements	2

This question was answered by 72 participants.

alpha = 0.79). High scores indicate low rejection of exclusion and low numbers indicate strong rejection of exclusion. Moreover, we asked the participants how likely they were to intervene, given the situation took place in their class. The likelihood of intervention was also assessed using a seven-point Likert-type scale (1 = very unlikely to 7 = very likely). Eventually, we asked the participants to justify their decision and to indicate what specific actions they would have taken (open-ended questions).

Coding of Open-Ended Questions

The coding systems for the open-ended questions are based on the study by Beißert and Bonefeld (2020) and were extended by inductively developing categories from the surveys themselves (see **Tables 1**, **2** for an overview and examples). The coding was completed by two independent coders, that were not allowed to code more than three relevant justifications for each statement. Based on 20% of the interviews we calculated a high interrater reliability, with Cohen's kappa = 0.96 for both, for the justifications of the likelihood of intervention as well as for the specific actions.

RESULTS

Data Analysis

Univariate ANOVAs were used to test for differences in the evaluation of exclusion and the likelihood of intervention between the different experimental conditions and between male and female participants.

Repeated measures ANOVAs on the proportional use of categories were conducted to analyze reasoning data from the open-ended questions. ANOVA frameworks are appropriate for repeated measures reasoning analyses because ANOVAs are robust to the problem of empty cells, whereas other data analytic procedures require cumbersome data manipulation to adjust for empty cells (see Posada and Wainryb, 2008, for a more thorough explanation and justification of this data analytic approach).

All analyses were firstly run with participants' age and years of service experiences included. But as there were no effects based on these variables in any of the analyses, we dropped these variables from the analyses for the sake of simpler models.

Evaluation of Exclusion

In line with our expectations, we found a general tendency to reject exclusion across both protagonists, i.e., a right-skewed distribution on the evaluation scale with a skewness of 0.21 (SE = 0.25), a mean of 2.94 (SD = 1.14), mode = 4.00, and median = 3.00.

To analyze differences in the evaluation of exclusion based on the gender of the participants and the gender of the excluded person, a 2 (participant gender: male, female) × 2 (protagonist gender: male, female) univariate ANOVA was conducted.

As expected, there was a main effect of participant gender, $F(1, 86) = 4.94$, $p = 0.029$, $\eta_p^2 = 0.05$, demonstrating that females (M = 2.81, SD = 1.01) rejected exclusion more strongly than males (M = 3.58, SD = 1.50). There was no effect of the gender of the protagonist, nor any interaction effects. See **Figure 1** for a graphical presentation of these results.

Likelihood of Intervention

This question was answered by 86 participants. The descriptive analyses showed that 56 participants (65.1%) tended to intervene, 22 participants (25.6%) tended not to intervene, and 8 (9.3%) participants chose the middle of the scale, indicating that it was as likely that they would intervene as not intervene.

To test for differences in the likelihood of intervention, a 2 (participant gender: male, female) × 2 (protagonist gender: male, female) univariate ANOVA was conducted. As preliminary analyses revealed no effects of participants' age or the years of participants' service experiences, these variables were not included in the analysis for the sake of a simpler model.

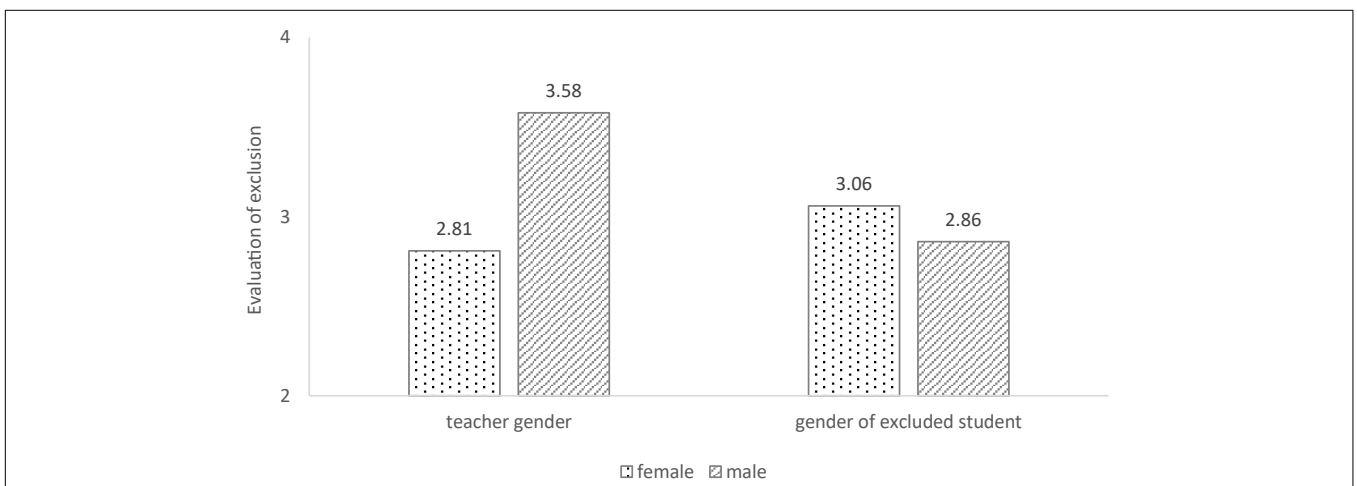


FIGURE 1 | Evaluation of exclusion as a function of teacher gender and the gender of the excluded student. High numbers indicate high acceptability of exclusion; low numbers indicate strong rejection of exclusion.

As expected, there was a main effect of the gender of the protagonist, $F(1, 82) = 14.11$, $p < 0.001$, $\eta_p^2 = 0.15$, revealing that participants were less likely to intervene in scenarios in that a boy was excluded ($M = 4.29$, $SD = 2.03$) compared to scenarios in that a girl was excluded ($M = 5.62$, SD = 1.37). There was no effect of the participants' gender, nor any interaction effects. See **Figure 2** for a graphical presentation of these results.

Justification of Likelihood of Intervention

To analyze the participants' justifications why they would tend to intervene or not as well as their specific actions, we conducted reasoning analyses on the proportional use of the coded categories. In order to see whether the specific justifications were related to the decision to intervene or not, we created a new variable out of the seven-point scale measuring the likelihood of intervention, resulting in the three categories "tendency to intervene," "indecisive," and "tendency not to intervene."

Using this new variable, we ran a 3 (decision: no intervention, indecisive, intervention) × 2 (participant gender: male, female) × 2 (protagonist: boy, girl) × 4 (justification: need for information, children's autonomy, empathy for the victim/avoid psychological harm, social norms of inclusion and cooperation) ANOVA with repeated measures on the factor "justification." The Huynh–Feldt adjustment was used to correct for violations of sphericity.

This analysis revealed an interaction effect of justification and decision, $F(6, 192) = 2.63$, $p < 0.018$, $\eta_p^2 = 0.08$, demonstrating that "need for information" and "social norms of inclusion and cooperation" were mainly used by those who tended to intervene, whereas "children's autonomy" was mainly used by those who tended not to intervene. However, there were no main or interaction effects based on the gender of the participants or the gender of the protagonist.

Specific Actions

In order to get a better understanding of *how* teachers would intervene, we analyzed their answers to the open-ended question of what they would specifically do when intervening in the situation. To test for differences in these answers based on the gender of the participants and the gender of the protagonist, a 2 (participant gender: male, female) × 2 (protagonist: boy, girl) × 4 (action: ask for reasons, conversation, inclusion-oriented behavior, find an alternative solution for excluded student) ANOVA was run with repeated measures on the factor "action." The Huynh–Feldt adjustment was used to correct for violations of sphericity.

This analysis revealed a main effect of action, $F(4.33, 44.23) = 6.63$, $p < 0.001$, $\eta_p^2 = 0.09$, indicating that participants stated significantly more often that they would ask for reasons or talk to the students than they would try to find an alternative solution for the excluded student. The frequency of inclusion-oriented behavior was not different from the other actions. Again, there were no main or interaction effects based on the gender of the participants or the gender of the protagonist.

DISCUSSION

The current study investigated teachers' reactions to social exclusion scenarios in Germany. Focusing on teachers as observers of social exclusion, we used hypothetical scenarios in which a student (girl vs. boy) was excluded by other children in class. We assessed teachers' evaluations of the exclusion behavior as well as how likely they were to intervene in the situation, and what they would specifically do. To extend prior research, we focused on the role of gender for teachers' reactions to social exclusion. More specifically, we focused on the gender of the excluded student on the one hand and the gender of the observing teacher on the other hand.

General Tendency to Reject Social Exclusion Among Students and to Intervene in Exclusion Situations Among Students

As expected, the teachers in our study showed a general tendency to reject social exclusion among students. This replicates the findings of other studies (e.g., Beißert and Bonefeld, 2020;

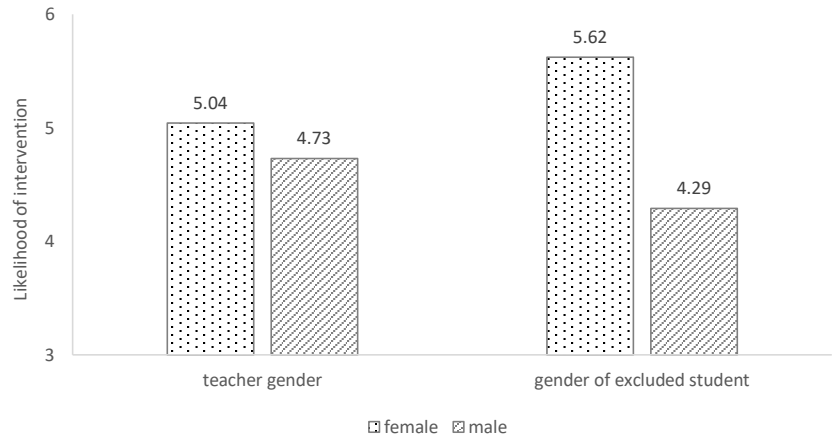


FIGURE 2 | Likelihood of intervention as a function of teacher gender and the gender of the excluded student. High numbers indicate high likelihood to intervene; low numbers indicate low likelihood to intervene.

Grütter et al., 2021; Kollerová and Killen, 2021), which overall provide strong evidence that teachers generally reject social exclusion among students. This tendency to reject exclusion seems to translate into action intentions insofar that the majority of the participating teachers stated that they would have intervened in the situation if it had happened in their class.

The Role of Teachers' Gender

Based on considerations of gender-specific socialization (Bakan, 1966; Cross and Madson, 1997; Zahn-Waxler, 2000) we had assumed that female teachers would reject social exclusion even more strongly than male teachers. In terms of the evaluation of social exclusion, we found evidence for this assumption. Interestingly, these gender differences in the evaluation of exclusion did not manifest in teachers' expected likelihood to intervene in the situation. Female teachers were not more likely to intervene in the situation than male teachers. One possible explanation for this discrepancy between the evaluation of and the expected reaction to social exclusion could also lie in gender-specific socialization. Women are socialized to connect with others and strive for companionship, but less to be self-effective agents (Bakan, 1966). In line with this, females attach great value to relationships (Guisinger and Blatt, 1994) and have a strong need for harmony (Hwang and Mattila, 2019). This might lead females to be hesitant to intervene in the exclusion situation because intervening might be conceptualized as getting involved in an interpersonal conflict. Hence, even though females reject exclusion more strongly, they might not take the step to action. However, one important limitation is that there were only 14% male teachers in the sample, and thus, the results should be considered with caution.

Interestingly, the low proportions of male participants are a typical problem of many online studies (Cull et al., 2005; Cheung et al., 2017; Beißert et al., 2020) and especially in studies on social exclusion (Butler, 2012; Butler and Shibaz, 2014; Beißert et al., 2021) with about 90% female participants. And even though we find a higher base rate of females compared to males among teachers, these samples as well as our sample include even more females than the proportion of female teachers in Germany (which would be appr. 73%, Statista Research Department, 2021). This might indicate that there is some self-selection of helpful individuals, since more helpful individuals are presumably more likely to participate in studies voluntarily. Nevertheless, further research should continue to investigate whether these results can be replicated in a sample with more male teachers.

The Role of the Excluded Student's Gender

The gender of the excluded student did not affect teachers' evaluation of the exclusion situation. The exclusion of boys and girls was evaluated equally reprehensible. However, the gender of the excluded student did influence teachers' behavioral intentions. As expected, teachers were more likely to intervene in scenarios in that a girl was excluded compared to scenarios in that a boy was excluded. This fits to our assumptions that this is due to gender expectations such as girls being more likely to strive to connect with others (communion) while boys tend to be more focused on individual goals (agency) (Bem, 1981). Therefore, boys may be seen as less vulnerable and less affected by social exclusion, which could induce a weaker need to intervene and protect them.

Justifications for Likelihood to Intervene and Specific Actions

We could not find any differences between female and male teachers' justifications to intervene or not, nor were there any differences based on the gender of the excluded student. Thus, the finding of Kollerová and Killen (2021) that teachers used more moral justifications when reasoning about excluded girls compared to excluded boys could not be replicated in the current study. Gender was not relevant for teachers' reasoning.

Interestingly, teachers' decisions to intervene or not were associated with different considerations. Namely, if teachers conceptualized the exclusion scenario as something that falls in the children's scope of action., i.e., when they referenced children's autonomy as a justification, they were less likely to intervene compared to those teachers who focused on socio-moral aspects such as inclusion and cooperation as social norms or compared to those who understood the situation to be ambiguous and stated their need for further information. That means that teachers' tendency to intervene seems to depend on how they perceive the exclusion situation—but independent of their own gender or the gender of the excluded student. In this context, it is encouraging that many teachers wanted to ask for reasons to better understand the situation in order to find out if further interventions were necessary or not.

Implications and Future Directions

All in all, we can say that gender is an important aspect in the context of social exclusion. On the one hand, the gender of the observing teacher is relevant as females reject exclusion even more strongly than males. On the other hand, the gender of the excluded student impacted teachers' reactions to social exclusion as they were less likely to intervene when a boy was excluded compared to a girl. We explain these findings with gender-specific socialization and social expectations linked to gender. Namely, the stronger focus on communal aspects in girls' socialization which is associated with a high value of relationships and harmony on the one hand, and the perception of females as being more vulnerable and more in need of relatedness than males, on the other hand. Future research should systematically examine whether such a communal orientation with a higher focus on interpersonal affiliation in females really can explain the current findings.

Additionally, further research should pay more attention to the assessment of evaluations and behavior or at least behavioral intentions. In this study, it becomes clear that even though the general tendency to reject exclusion among students manifests in a general tendency to intervene, the effects related to gender reveal differential patterns regarding the evaluations and the

behavioral tendency. Hence, further research should include both attitudinal and behavioral measures. Moreover, in this context, it would be of great interest to conduct real behavioral studies in naturalistic settings in order to investigate whether the reported behavioral intentions transmit into the respective actions.

Interestingly, the gender differences that we found regarding the reported behavioral intentions are not reflected in teachers' reasoning or their specific actions. That means, teachers are more likely to intervene when girls are excluded than when boys are excluded. However, once they decide to intervene, the specific actions are not related to the gender of the excluded student. This leads us to the assumption that the differences in the likelihood to intervene or not are no conscious tendencies but rather automatisms based on socialized expectations linked to gender. Thus, it is crucial to sensitize teachers to such expectations and help them reflect their own gender-specific expectations. Further, teachers should be encouraged to treat both genders equally and to consequently intervene also when boys are excluded. It is important to make teachers aware of the fact that boys and girls suffer equally from the severe consequences of social exclusion.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HB and MB developed the idea and the design of the study. MS made a first draft of the manuscript which was revised by HB and MB. All authors have approved the final version to be published and agreed to be accountable for all aspects of the work and ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved, and have contributed meaningfully to the manuscript and analyzed and interpreted the data.

ACKNOWLEDGMENTS

We would like to thank the undergraduate research assistants Michelle Barth and Julia Hildebrandt for their help to prepare and conduct this study and Lea Markhoff for her support with the data preparation and literature research. And we are grateful to all teachers who participated in this study.

REFERENCES

Bakan, D. (1966). *The Duality of Human Existence*. Chicago: Rand McNally.

Beißert, H., and Bonefeld, M. (2020). German Pre-service Teachers' Evaluations of and Reactions to Interethnic Social Exclusion Scenarios. *Front. Educ.* 5:586962. doi: 10.3389/feduc.2020.586962

Beißert, H., Köhler, M., Rempel, M., and Kruyen, P. (2020). Ein Vergleich traditioneller und computergestützter Methoden zur Erstellung einer deutschsprachigen Need for Cognition Kurzskala. *Diagnostica* 66, 37–49. doi: 10.1026/0012-1924/a000242

Beißert, H., Staat, M., and Bonefeld, M. (2021). The role of ethnic origin and situational information in teachers' reactions to social exclusion among students. *Soc. Psychol. Educ.* 24, 1511–1533. doi: 10.1007/s11218-021-09656-5

Beißert, H., Bonefeld, M., Schulz, I., and Studenroth, K. (2019). *Beurteilung sozialer Ausgrenzung unter Erwachsenen: Die Rolle des Geschlechts. [Poster presentation]*. Mannheim: Moraltagung.

Bem, S. L. (1981). Gender schema theory: a cognitive account of sex typing. *Psychol. Rev.* 88, 354–364. doi: 10.1037/0033-295X.88.4.354

Bonefeld, M., and Dickhäuser, O. (2018). (Biased) Grading of Students' Performance: students' Names, Performance Level, and Implicit Attitudes. *Front. Psychol.* 9:481. doi: 10.3389/fpsyg.2018.00481

Buhs, E. S., Ladd, G. W., and Herald, S. L. (2006). Peer exclusion and victimization: processes that mediate the relation between peer group rejection and children's classroom engagement and achievement? *J. Educ. Psychol.* 98, 1–13. doi: 10.1037/0022-0663.98.1.1

Butler, R. (2012). Striving to connect: extending an achievement goal approach to teacher motivation to include relational goals for teaching. *J. Educ. Psychol.* 104, 726–742. doi: 10.1037/a0028613

Butler, R., and Shibaz, L. (2014). Striving to connect and striving to learn: influences of relational and mastery goals for teaching on teacher behaviors and student interest and help seeking. *Int. J. Educ. Res.* 65, 41–53. doi: 10.1016/j.ijer.2013.09.006

Cheung, K. L., Ten Klooster, P. M., Smit, C., de Vries, H., and Pieterse, M. E. (2017). The impact of non-response bias due to sampling in public health studies: a comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health* 17:276. doi: 10.1186/s12889-017-4189-8

Cross, S. E., and Madson, L. (1997). Models of the self: self-construals and gender. *Psychol. Bull.* 122, 5–37. doi: 10.1037/0033-2909.122.1.5

Cull, W. L., O'Connor, K. G., Sharp, S., and Tang, S. F. S (2005). Response rates and response bias for 50 surveys of pediatricians. *Health Services Res.* 40, 213–226. doi: 10.1111/j.1475-6773.2005.00350.x

D'Ambrosio, F., Olivier, M., Didon, D., and Besche, C. (2009). The basic empathy scale: a French validation of a measure of empathy in youth. *Pers. Individ. Dif.* 46, 160–165. doi: 10.1016/j.paid.2008.09.020

Eckes, T. (2010). "Geschlechterstereotype: von rollen, identitäten und vorurteilen," in *Handbuch Frauen- und Geschlechterforschung. Theorie, Methoden*, Empirie, eds R. Becker and B. Kortendiek (Wiesbaden: VS Verlag für Sozialwissenschaften), 178–189.

Ely, R., Melzi, G., Hadge, L., and McCabe, A. (1998). Being brave, being nice: themes of agency and communion in children's narratives. *J. Pers.* 66, 257–284. doi: 10.1111/1467-6494.00012

Fuhrmann, D., Casey, C. S., Speekenbrink, M., and Blakemore, S.-J. (2019). Social exclusion affects working memory performance in young adolescent girls. *Dev. Cogn. Neurosci.* 40:100718. doi: 10.1016/j.dcn.2019.100718

Gazelle, H., and Druhen, M. J. (2009). Anxious solitude and peer exclusion predict social helplessness, upset affect, and vagal regulation in response to behavioral rejection by a friend. *Dev. Psychol.* 45, 1077–1096. doi: 10.1037/a0016165

Gilligan, C. (1993). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.

Grütter, J., Barth, C., Tschopp, and Buholzer, A. (2021). "Teachers' emotional reactions and reasoning about social exclusion based on learning difficulties and hyperactive behavior," in *Proceedings of the 47th AME Conference.*

Guisinger, S., and Blatt, S. J. (1994). Individuality and relatedness: evolution of a fundamental dialectic. *Am. Psychol.* 49, 104–111. doi: 10.1037/0003-066X.49.2.104

Horn, S. S. (2003). Adolescents' reasoning about exclusion from social groups. *Dev. Psychol.* 39, 71–84. doi: 10.1037/0012-1649.39.1.71

Hwang, Y., and Mattila, A. S. (2019). Feeling left out and losing control: the interactive effect of social exclusion and gender on brand attitude. *Int. J. Hosp. Manag.* 77, 303–310. doi: 10.1016/j.ijhm.2018.07.010

Jiang, S., and Ngai, S. S.-Y. (2020). Social exclusion and multi-domain well-being in Chinese migrant children: exploring the psychosocial mechanisms of need satisfaction and need frustration. *Child. Youth Serv. Rev.* 116:105182. doi: 10.1016/j.childyouth.2020.105182

Killen, M., Mulvey, K. L., and Hitti, A. (2013). Social exclusion in childhood: a developmental intergroup perspective. *Child Dev.* 84, 772–790. doi: 10.1111/cdev.12012

Killen, M., and Stangor, C. (2001). Children's social reasoning about inclusion and exclusion in gender and race peer group contexts. *Child Dev.* 72, 174–186. doi: 10.1111/1467-8624.00272

Kollerová, L., and Killen, M. (2021). An experimental study of teachers' evaluations regarding peer exclusion in the classroom. *Br. J. Educ. Psychol.* 91, 463–481. doi: 10.1111/bjep.12373

Malti, T., Killen, M., and Gasser, L. (2012). Social judgments and emotion attributions about exclusion in Switzerland. *Child Dev.* 83, 697–711. doi: 10.1111/j.1467-8624.2011.01705.x

Mello, S., Tan, A. S. L., Sanders-Jackson, A., and Bigman, C. A. (2019). Gender Stereotypes and Preconception Health: men's and Women's Expectations of Responsibility and Intentions to Engage in Preventive Behaviors. *Matern. Child Health J.* 23, 459–469. doi: 10.1007/s10995-018-2654-3

Mulvey, K. L., Gönültaş, S., Irdam, G., Carlson, R. G., DiStefano, C., and Irvin, M. J. (2021). School and Teacher Factors That Promote Adolescents' Bystander Responses to Social Exclusion. *Front. Psychol.* 11:581089. doi: 10.3389/fpsyg.2020.581089

Murray-Close, D., and Ostrov, J. M. (2009). A longitudinal study of forms and functions of aggressive behavior in early childhood. *Child Dev.* 80, 828–842. doi: 10.1111/j.1467-8624.2009.01300.x

Neuburger, S., Ruthsatz, V., Jansen, P., and Quaiser-Pohl, C. (2015). Can girls think spatially? Influence of implicit gender stereotype activation and rotational axis on fourth graders' mental-rotation performance. *Learn. Individ. Differ.* 37, 169–175. doi: 10.1016/j.lindif.2014.09.003

Posada, R., and Wainryb, C. (2008). Moral development in a violent society: colombian children's judgments in the context of survival and revenge. *Child Dev.* 79, 882–898. doi: 10.1111/j.1467-8624.2008.01165.x

Retelsdorf, J., Schwartz, K., and Asbrock, F. (2015). Michael can't read! *J. Educ. Psychol.* 107, 186–194. doi: 10.1037/a0037107

Riva, P., and Eck, J. (2016). "The Many Faces of Social Exclusion," in *Social Exclusion: Psychological Approaches to Understanding and Reducing its Impact*, 1st Edn, eds J. Eck and P. Riva (Cham: Springer), ix–xv.

Rueckert, L., and Naybar, N. (2008). Gender differences in empathy: The role of the right hemisphere. *Brain Cogn.* 67, 162–167. doi: 10.1016/j.bandc.2008.01.002

Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., and Piefke, M. (2008). Gender differences in brain networks supporting empathy. *Neuroimage* 42, 393–403. doi: 10.1016/j.neuroimage.2008.04.180

Sebastian, C., Viding, E., Williams, K. D., and Blakemore, S.-J. (2010). Social brain development and the affective consequences of ostracism in adolescence. *Brain Cogn.* 72, 134–145. doi: 10.1016/j.bandc.2009.06.008

Smetana, J. G. (1989). Toddlers' social interactions in the context of moral and conventional transgressions in the home. *Dev. Psychol.* 25, 499–508. doi: 10.1037/0012-1649.25.4.499

Spence, J. T., and Helmreich, R. L. (1978). *Masculinity and Femininity: Their Psychological Dimensions, Correlates, and Antecedents*. Austin, TX: University of Texas Press.

Statista Research Department (2021). *Frauenanteil unter den Lehrkräften nach Schulart 2021*. Available online at: https://de.statista.com/statistik/daten/studie/1129852/umfrage/frauenanteil-unter-den-lehrkraeften-in-deutschland-nach-schulart/ (accessed November 15, 2021).

Stuijfzand, S., Wied, M., de Kempes, M., van de Graaff, J., Branje, S., and Meeus, W. (2016). Gender Differences in Empathic Sadness towards Persons of the Same- versus Other-sex during Adolescence. *Sex Roles* 75, 434–446. doi: 10.1007/s11199-016-0649-3

Szekely, L., Bonefeld, M., and Beißert, H. (under review). *Teachers' Reactions to Social Exclusion among Students: The Effect of Ethnic Origin and Background Information*. Open Psychology.

Tay, P. K. C., Ting, Y. Y., and Tan, K. Y. (2019). Sex and care: the evolutionary psychological explanations for sex differences in formal care occupations. *Front. Psychol.* 10:867. doi: 10.3389/fpsyg.2019.00867

Van der Graaff, J., Branje, S., Wied, M., de Hawk, S., van Lier, P., and Meeus, W. (2014). Perspective taking and empathic concern in adolescence: gender differences in developmental changes. *Dev. Psychol.* 50, 881–888. doi: 10.1037/a0034325

Wesselmann, E. D., Grzybowski, M. R., Steakley-Freeman, D. M., DeSouza, E. R., Nezlek, J. B., and Williams, K. D. (2016). "Social exclusion in everyday life," in *Social Exclusion: Psychological Approaches to Understanding and Reducing its Impact*, 1st Edn, eds J. Eck and P. Riva (Cham: Springer), 323.

Wesselmann, E. D., Williams, K. D., and Hales, A. H. (2013). Vicarious ostracism. *Front. Human Neurosci.* 7:153. doi: 10.3389/fnhum.2013.00153

Williams, K. D. (2009). "Chapter 6 Ostracism," in *Advances in Experimental Social Psychology*, Vol. Vol. 41, ed. M. P. Zanna (Cambridge, US: Academic Press), 275–314. doi: 10.1016/S0065-2601(08)00406-1

Zahn-Waxler, C. (2000). "The development of empathy, guilt, and internalization of distress: Implications for gender differences in internalizing and externalizing problems," in *Wisconsin Symposium on Emotion: Anxiety, Depression, and Emotion*, ed. R. Davidson (Oxford, UK: Oxford University Press), 222–265.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 14 February 2022
doi: 10.3389/fpsyg.2021.813572

Mothers' and Fathers' Science-Related Talk With Daughters and Sons While Reading Life and Physical Science Books

*Tess A. Shirefley and Campbell Leaper**

Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA, United States

Introduction: In prior studies conducted in the United States, parents' gender-differentiated encouragement of science predicted children's later science motivation. Most of this research has focused on older children or teens and only looked at the impact of mothers. However, accumulating evidence suggests that gender-differentiated encouragement of science interest may begin in early childhood. Moreover, fathers may be more likely than mothers to treat sons and daughters differently in science-learning contexts.

Methods: We examined 50 United States families with both a mother and a father (82% White; 98% with at least some college education) and either a daughter or a son (48–83 months; $M = 62$, $SD = 9$). On separate visits, each parent reads two books with their child. One was about life science and the other was about physical science. We coded parents' science-related talk during these interactions.

Results and Conclusion: In contrast to our predictions, parents used higher proportions of science talk with daughters than sons, including higher average rates of overall science talk and specific types of science talk (e.g., science explanations, science-related personal connections, and science-learning talk). Moreover, most of the child gender effects occurred while reading the physical science books. Book topic and parent gender moderated some additional patterns. Book reading is discussed as a potential context for mitigating socialization experiences that traditionally disfavor girls' interest in physical science.

Keywords: gender differences, mother-child communication, father-child communication, reading, science education

OPEN ACCESS

Edited by:
Joyce Endendijk, Utrecht University, Netherlands

Reviewed by:
Katherine Short-Meyerson, University of Wisconsin–Oshkosh, United States
Liesbeth Hallers-Haalboom, Utrecht University, Netherlands
Kingsley Schroeder, Springfield College, United States

***Correspondence:**
Campbell Leaper
cam@ucsc.edu

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *12 November 2021*
Accepted: *29 December 2021*
Published: *14 February 2022*

Citation:
Shirefley TA and Leaper C (2022) Mothers' and Fathers' Science-Related Talk With Daughters and Sons While Reading Life and Physical Science Books. Front. Psychol. 12:813572. doi: 10.3389/fpsyg.2021.813572

INTRODUCTION

Even though women and men demonstrate comparable levels of participation in the life science workforce in the United States, women remain underrepresented in physical science domains (National Science Foundation, 2021). As documented in earlier reports, interest in physical science more likely increased from middle childhood to adolescence among boys than girls; in contrast, interest in life science remained comparable for girls and boys during this period (e.g., Baram-Tsabari and Yarden, 2008). The development of average gender differences in motivation and

achievement is attributed to a combination of individual, interpersonal, and cultural factors (Cheryan et al., 2017; Eccles and Wigfield, 2020). Among them, researchers have highlighted the potential impact of parents' gender-differentiated socialization on children's developing interests and ability beliefs (Eccles and Wigfield, 2020). We built on prior research in three ways. First, previous research on parents' gender-differentiated socialization of their children's science learning and interest has focused on middle childhood and adolescence. Hence, we explored whether the gender-differentiated patterns might be detected in a younger age group of children between 4 and 7 years of age. Second, previous research studies looked primarily at mothers without considering parent gender as a potential moderator. We examined children's book reading separately with their mothers and fathers. Finally, scant research has separately examined physical science and life science when considering parents' gender-differentiated treatment. Hence, we observed parents reading separate science books on life science and physical science with their children.

Parents' Gender-Differentiated Socialization of Children's Science Interest

Longitudinal studies established how parents' gender-differentiated beliefs about their children's science, technology, engineering, and mathematics (STEM)-related interests or abilities predicted later changes in children's motivational beliefs and achievement (e.g., Simpkins et al., 2015). However, if parents' gender-stereotyped expectations matter, then how are they manifested in their interactions with their children at young ages? According to ecological and social cognitive theories of development (Bussey and Bandura, 1999; Bronfenbrenner, 2005), this can occur when parents provide different opportunities for learning to children based on their gender. For example, this was indicated by Tenenbaum and Leaper (2003) in their observations of parents with their 10-year-old daughter or son while engaging in assigned science activities. On average, fathers (but not mothers) used more science-related talk (e.g., explanations, scientific vocabulary) with their sons than daughters during a physical science task. But other studies suggest that this kind of gender-differentiated treatment may occur at much younger ages. In at least three studies, parents of preschool-age children were observed talking more about science with their sons than daughters. These effects were observed at a science museum (Crowley et al., 2001) while playing with a physics toy at home (Tenenbaum et al., 2005) and reading a science-related book (Shirefley et al., 2020). Although the evidence is limited, two studies suggest that gender-related variations might occur with some types of science talk more than others (Tenenbaum et al., 2005; Shirefley et al., 2020).

Shared Book Reading as a Context for Investigating Parents' Talk With Young Children

Shared book reading is a common context in many families where informal learning for young children occurs with their parents (Scholastic Inc, 2016). More specifically, researchers have highlighted how parents' book reading with preschool-age children was a means for discussing and learning complex science concepts (e.g., Kelemen et al., 2014; Shirefley et al., 2020) and imparting lessons about gender roles (e.g., Friedman et al., 2007; Endendijk et al., 2014). However, no prior work has considered how conversations during shared reading may vary with the type of science book (life vs. physical) or with both mothers and fathers.

Comparing Fathers' and Mothers' Science Talk With Children

When considering parents' gender as a moderator of science talk, two patterns have been previously identified. First, average differences between mothers' and fathers' behavior with their children may occur (refer to Leaper, 2015 for review). Some studies have found that mothers were more verbal than fathers when interacting with children (refer to Leaper et al., 1998 for a meta-analysis). Only a few studies have compared mothers' and fathers' verbal behavior specifically during shared reading with their young children. Their results have been mixed. Two studies observed greater talking or more teaching-related comments among mothers than fathers (Conner et al., 1997; Schwartz, 2004). One study noted more talking among fathers than mothers (Anderson et al., 2004). In addition, another investigation found negligible differences between mothers' and fathers' teaching-related speech during shared reading (Blake et al., 2006). None of these studies, however, observed the shared reading of books focused on science topics.

A second pattern regarding parent gender differences indicated in the research literature is for fathers to be more likely than mothers to treat daughters and sons differently (refer to Leaper, 2015). Regarding science-related talk with young children, one study observed that both mothers and fathers used more science talk with boys than girls at a science museum, but the trend was stronger among fathers (Crowley et al., 2001). We do not know whether similar patterns would be seen while reading science books.

In addition, research with older children suggests that differences between fathers and mothers in gender-differentiated encouragement of science may partly depend on the type of science. Two studies looked at mothers' and fathers' hands-on involvement with elementary-school-age children in both life science and physical science tasks. In the first study, fathers used more science-teaching talk with sons than daughters but only during the physical science task; conversely, mothers did not differ with daughters and sons in either task (Tenenbaum and Leaper, 2003). In the second study, researchers surveyed parents based on the kinds of science problems they solved with their children (Short-Meyerson et al., 2016). Mothers favored more life science tasks, whereas fathers preferred more physical science tasks. From these studies, there is evidence to suggest that average differences in mothers' and fathers' behavior may occur when reading science books to their young children. Accordingly, we took into account the parent gender, the child gender, and the type of science book being read.

Current Study

To build on earlier research investigating parents' science-related talk with children, we observed parents with their 4–7-year-old children while reading physical and life science books in their homes. We chose book reading as it is a common shared activity among many parents and their young children, and it is an activity that is easily arranged in families' homes. We tested for variations in parents' science-related talk by child gender, parent gender, and the type of science book. Our hypotheses were as follows: first, we expected that parents would use a greater proportion of science-related talk with their sons than daughters. Second, we predicted that parents' gender-differentiated science talk would be more likely for fathers than mothers. Finally, we hypothesized that these effects would be stronger when reading the physical science book. When conducting our analyses, we looked at parents' overall science talk. In addition, we examined specific forms of science talk to explore whether some might be related to gender-related variations more than others. Among the few studies that examined parents' science talk, none of them considered whether gender-differentiated treatment was more likely for some forms of science talk than others.

METHODS

Participants

Participants were recruited in northern California through social media posts, local community spaces, and preschools. This study focuses on families in our sample with heterosexual parents in which both the mother and the father were able to participate, which initially comprised 55 families. Of these families, five were removed due to technical difficulties ($n = 2$) or child non-compliance with the tasks ($n = 3$). Our analyses are based on 50 families with a participating daughter or son ($n = 25$ each) between 4 and 7 years of age ($M = 62$ months, $SD = 9.5$). The average age of daughters and sons did not significantly differ. For mothers, 82% self-identified as White and 88% had attained at least a bachelor's degree. For fathers, 82% self-identified as White and 74% had attained at least a bachelor's degree (refer to **Table 1** for more detail). Parent-child dyads were asked to read and discuss the books as it was most natural to them ($n = 48$ exclusively in English, $n = 1$ in English and Spanish, and $n = 1$ in English and German).

MATERIALS AND PROCEDURE

Science Books

We selected four science books from the *Let's Read-and-Find-Out Science* series by HarperCollins written for preschool-aged children. All books were in English. Two different books focused on physical science [*What Is The World Made Of?* (about solids, liquids, gasses) and *Light Is All Around Us* (how light brightens the world)] and two books were on life science [*From Seed to Pumpkin* (process of a pumpkin growing) and *From Caterpillar To Butterfly* (transformation from caterpillar to butterfly)]. We edited the four books to balance the proportion of science-related content across books. Each book was about 13 pages with approximately 30% of the text containing science words.

After obtaining signed consent, parent-child dyads were video-recorded while reading one life science book and one physical science book (with different versions provided to mothers and fathers). No time limit was imposed. The order of books and the versions of each book type were counterbalanced across parent gender and child gender. We attempted to counterbalance the order of visits with mothers and fathers; however, several fathers would not participate in the study unless mothers participated first. Mothers were visited first in 17 of 25 families with daughters and 16 of 25 families with sons. Home visits were conducted 1–2 weeks apart.

After parents completed the reading of the science books, they completed a brief survey assessing their attitudes and beliefs about science. Among these questions, two items assessed their beliefs about their child's science ability and interest (from Tenenbaum and Leaper, 2003): "My child finds science ($0 =$ *very boring* to $7 =$ *very interesting*)" and "My child finds science…($0 =$ *very hard* to $7 =$ *very easy*)." Also, we asked "How often do you read a storybook to your child?" ($0 =$ *never*, $1 =$ *a few times per year*, $2 =$ *about once per month*, $3 =$ *about once per week*, and $4 =$ *almost every day*).

Coding

We first transcribed parent-child video recordings using Datavyu. We parsed parent-child talk into utterances representing individual thought units. Parent and child utterances were coded into 16 coding categories, which included five science-related codes (based on Shirefley et al., 2020). Five research assistants coded 20% of the dataset to assess intercoder reliability. After achieving reliability and discussing differences, each coder coded 20% of the remaining samples. The five types of the science-related talk were as follows: *scientific explanations*, *science labels*, *scientific personal connections*, *scientific story inferences*, and *scientific-learning talk* (refer to **Table 2** for definitions and **Table 3** for descriptive statistics). Based on

TABLE 1 | Demographic backgrounds of mothers and fathers.

Variable	Mothers	Fathers
Ethnicity		
White	41	41
Latinx	5	6
Asian/Pacific Islander	1	2
Black	1	0
Multi/Other	2	1
Education level		
High school diploma	1	1
Some college/Associate's	5	12
College bachelors	19	20
Masters/Doctorate/Medical	25	17

the guidelines developed by Landis and Koch (1977), the intercoder agreement for each code was acceptable (refer to **Table 2**).

RESULTS

Statistical Design

Children were observed on separate occasions with their mother and father, and each parent reads two types of science books. Because members of dyads are not independent, we utilized linear mixed models to conduct our analyses (Kenny et al., 2006). Child gender was a between-group factor, whereas parent gender and book type were nested factors. The mixed linear model is only able to examine one criterion variable at a time. Accordingly, we ran six models with parents' overall science talk and the five specific types of science talk. To control for variations across parents in the time spent talking about the book, we calculated each type of science talk as a proportion of total utterances (excluding reading text from the book).

Preliminary Analyses

We conducted preliminary analyses to test for gender-related variations in a few factors that might influence parents' gender-differentiated talk. First, we did not find significant differences in parents' views of daughters' and sons' science ability or interest, although mothers were more likely than fathers to rate their children as finding science easy ($p = 0.032$). Second, we did not find differences in parents' reported reading to daughters vs. sons. Indeed, 94% of mothers and 90% of fathers reported reading to their children "almost every day." Finally, we did not find differences between daughters' and sons' total talk and science-related talk with either science book, although children used proportionally more science talk with the life science book than the physical science book ($p < 0.001$).

Testing Hypotheses

In summarizing the results below, only significant effects from the models are noted (refer to **Table 4** for more information). With any significant pairwise comparisons tests, Cohen's *d* indices of

TABLE 2 | Science talk codes: descriptions and intercoder reliability.

Measure	Definition	Percent agreement	Kappa coefficient	Evaluation[1]
Scientific explanations	Generic facts vocabulary and explanations about a phenomenon specifically related to the scientific material (e.g., "Roots suck up water like a straw").	88	0.66	Substantial
Science labels	The naming of a specific part of an image within the book (e.g., "Those are called pupa").	95	0.65	Substantial
Scientific personal connections	Relating the scientific material of the book to a child/parent/family's prior experience (e.g., "Remember when we made Play-Doh and at first it was really liquidy but then we added more starch to make it solid?").	95	0.72	Substantial
Scientific story inferences	Anticipating the next step in the story (taking information not visible on the page of the book to then infer understanding (e.g., "There was a lot of rain, I wonder what will happen to the pumpkin seeds").	99	0.37	Fair
Scientific-learning talk	A reference to new scientific knowledge gained or the opportunity for parent or child to check in with each other about their understanding of information (e.g., "Did you know the sun was that hot?!").	96	0.42	Moderate

[1] *Source: Landis and Koch (1977).*
*Story inferences occurred infrequently (refer to **Table 3**), which likely accounts for the relatively low intercoder agreement.*

TABLE 3 | Mean frequencies and proportions for science-related talk variables of parents.

	Frequencies		Proportions	
	Life science *M* (*SD*)	Physical science *M* (*SD*)	Life science *M* (*SD*)	Physical science *M* (*SD*)
Total talk	50.8 (25.5)	53.3 (36.7)	N/A	N/A
Overall science talk	30.8 (17.8)	37.8 (24.4)	0.60 (0.13)	0.59 (0.14)
Scientific explanations or vocabulary	13.0 (10.5)	22.6 (18.1)	0.24 (0.12)	0.34 (0.14)
Science labels	5.7 (4.3)	3.7 (4.4)	0.12 (0.09)	0.06 (0.06)
Scientific personal connections	6.7 (5.5)	6.7 (6.2)	0.14 (0.09)	0.11 (0.08)
Scientific story inferences	0.8 (2.1)	0.3 (0.6)	0.01 (0.03)	0.004 (0.01)
Scientific-learning talk	4.5 (3.8)	4.9 (4.1)	0.09 (0.05)	0.08 (0.05)

N/A, not applicable.
The total talk reflects all utterances excluding reading text from the book. Proportion scores reflect the proportions of each science talk variable in relation to parents' total utterances (excluding reading text from book).

TABLE 4 | Summary of results from linear mixed models.

	Overall $F_{(1,46)}$	Explanations $F_{(1,47)}$	Labels $F_{(1,48)}$	Connections $F_{(1,48)}$	Inferences $F_{(1,47)}$	Learning $F_{(1,47)}$
Child gender (CG)	4.70*	4.14*	0.07	0.27	0.17	8.98**
Parent gender (PG)	1.46	1.64	0.63	0.64	2.42	0.65
Book type (BT)	0.38	63.57***	44.75***	9.62**	11.92***	2.14
CG × BT	5.24*	1.03	0.01	7.33*	0.29	7.15**
PG × BT	0.32	1.15	0.05	2.07	6.93*	1.25
CG × PG	0.13	0.02	0.01	1.61	0.16	8.36**
CG × PG × BT	0.33	0.43	0.03	0.33	0.04	0.18

*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

effect size are reported. Effect sizes are negligible when $d < 0.2$ (or $\eta^2 < 0.01$), small when $d = 0.2$ (or $\eta^2 = 0.01$), moderate when $d = 0.5$ (or $\eta^2 = 0.06$), and large when $d = 0.8$ (or $\eta^2 = 0.14$) or greater (Cohen, 1988).

The Proportion of Overall Science Talk

The main effect of child gender occurred whereby parents used more overall science talk with daughters ($M = 0.62$, $SD = 0.12$) than sons ($M = 0.57$, $SD = 0.15$), $F_{(1,46)} = 4.70$, $p = 0.035$, $\eta^2_{partial} = 0.09$, $d = 0.37$. This main effect was subsumed by a Science Topic × Child Gender interaction, $F_{(1, 46)} = 4.50$, $p = 0.040$, $\eta^2_{partial} = 0.09$. Follow-up pairwise comparisons revealed that parents reading the physical science book used a higher average proportion of overall science talk with daughters ($M = 0.64$, $SD = 0.13$) than sons ($M = 0.55$, $SD = 0.14$), $p = 0.003$, $d = 0.67$. Parents did not significantly differ in their overall science talk with daughters ($M = 0.62$, $SD = 0.11$) and sons ($M = 0.59$, $SD = 0.16$) when reading the life science book.

The Proportion of Science Explanations

A significant main effect of child gender indicated that parents were more likely to use science explanations with daughters ($M = 0.31$, $SD = 0.15$) than sons ($M = 0.26$, $SD = 0.14$), $F_{(1, 47)} = 4.60$, $p = 0.037$, $\eta^2_{partial} = 0.09$. Also, a significant main effect of science topic revealed that parents were more likely to use science explanations when reading the physical science book ($M = 0.34$, $SD = 0.15$) than the life science book ($M = 0.22$, $SD = 0.12$), $F_{(1,47)} = 63.57$, $p = 0.003$, $\eta^2_{partial} = 0.18$.

The Proportion of Science Labeling

A main effect of science topic indicated that parents used proportionally more scientific labels on average when reading the life science book ($M = 0.12$, $SD = 0.09$) than the physical science book ($M = 0.06$, $SD = 0.06$), $F_{(1, 48)} = 44.75$, $p < 0.001$, $\eta^2_{partial} = 0.10$.

The Proportion of Science Personal Connections

Based on the main effect of science books, parents were more likely to make science-related personal connections when reading the life science book ($M = 0.14$, $SD = 0.09$) than the physical science book ($M = 0.10$, $SD = 0.08$), $F_{(1,48)} = 9.62$, $p = 0.001$, $\eta^2_{partial} = 0.20$. In addition, there was a significant Science Topic × Child Gender interaction, $F_{(1,48)} = 6.10$, $p = 0.017$, $\eta^2_{partial} = 0.12$. Follow-up pairwise comparison tests revealed child gender differences based on the science topic. When reading the physical science book, parents made more science-related personal connections with daughters ($M = 0.13$, $SD = 0.09$) than sons ($M = 0.08$, $SD = 0.05$), $p = 0.04$, $d = 0.69$. When reading the life science book, parents did not significantly differ in their use of scientific personal connections with daughters ($M = 0.13$, $SD = 0.08$) and sons ($M = 0.15$, $SD = 0.11$).

The Proportion of Science Inferences

The main effect of science topic revealed that parents used significantly more science inferences when reading the life science book ($M = 0.01$, $SD = 0.03$) than the physical science book ($M = 0.00$, $SD = 0.01$), $F_{(1,47)} = 11.92$, $p = 0.003$, $\eta^2_{partial} = 0.17$. A significant Parent Gender × Science Topic interaction [$F_{(1,47)} = 6.9$, $p = 0.011$, $\eta^2_{partial} = 0.13$] indicated a significant parent gender difference depending on science topic. On average, fathers ($M = 0.02$, $SD = 0.04$) were more likely than mothers ($M = 0.01$, $SD = 0.02$) to use science inferences when reading the life science book, $p = 0.012$, $d = 0.47$; but mothers ($M = 0.01$, $SD = 0.01$) and fathers ($M = 0.00$, $SD = 0.06$) did not significantly differ when reading the physical science book.

The Proportion of Science-Learning Talk

A significant main effect of child gender showed that parents generally used more science-learning talk with daughters than sons, $F_{(1,47)} = 10.5$, $p = 0.002$, $\eta^2_{partial} = 0.19$. However, this effect was subsumed into two interaction effects. First, there was a Child Gender × Parent Gender interaction, $F_{(1,47)} = 9.6$, $p = 0.003$, $\eta^2_{partial} = 0.17$. Follow-up pairwise comparison tests revealed that mothers used proportionally more science-learning talk with daughters ($M = 0.11$, $SD = 0.06$) than sons ($M = 0.05$, $SD = 0.04$), $p < 0.001$, $d = 1.08$. There was no significant difference in fathers' science-learning talk with daughters ($M = 0.08$, $SD = 0.05$) and sons ($M = 0.08$, $SD = 0.05$). Also, a significant Child Gender × Science Topic interaction occurred, $F_{(1,47)} = 9.5$, $p = 0.003$, $\eta^2_{partial} = 0.17$. On average, while reading the life science book, parents used a higher proportion of science-learning talk with daughters ($M = 0.11$, $SD = 0.06$) than sons ($M = 0.06$, $SD = 0.05$),

$p < 0.001$, $d = 0.91$. There were no significant differences in science-learning talk used between daughters ($M = 0.08$, $SD = 0.06$) and sons ($M = 0.07$, $SD = 0.05$) when reading the physical science book.

DISCUSSION

Our findings revealed patterns of gender differentiation in the science talk of mothers and fathers when reading physical and life science books to their sons or daughters. In contrast to our hypothesis, parents used a higher proportion of several forms of science talk with daughters compared to sons. Moreover, the magnitudes of these average differences were moderate-to-large in size. The science topic moderated some of these differences and parent gender moderated child gender differences although not always in an expected manner. Also, gender-related differences occurred across various types of science talk. As discussed below, the results suggested that science book reading with young children may be a context in some families in which parents may especially engage girls in science learning.

To the best of our knowledge, only three prior studies examined gender differences in parents' science talk with preschool- and early elementary-aged children. In these investigations, parents were more likely to use science explanations with boys than girls at a science museum (Crowley et al., 2001) while playing with a science activity at home (Tenenbaum et al., 2005) or reading a science-related book (Shirefley et al., 2020). We observed the opposite pattern whereby parents generally used more scientific explanations and overall science talk with their daughters than sons across both books. In addition, mothers (but not fathers) used more science-learning talk with daughters than sons across both books.

The science topic moderated some additional effects. When reading the physical science book, parents used proportionally more overall science talk and made more science-related personal connections with their daughters than sons. Given the gender gap in motivation and achievement in the physical sciences observed during adolescence, i.e., when boys have often participated in the physical sciences than girls (refer to Cheryan et al., 2017), this pattern was surprising.

Our sample comprised mothers and fathers who generally were highly educated and active readers with their children. Perhaps these parents made concerted efforts to counteract cultural stereotypes about gender and science (e.g., physical sciences being stereotypically masculine). In doing so, they may have sought to engage their daughters especially in the science topic that was most counter-stereotypical. Mothers, in particular, may have been focused on this goal, as we found mothers but not fathers used more science-learning talk with daughters than sons. Families in our study also lived near many scientific/technology industries and in communities where issues of gender and STEM are often highlighted in local and national media. In one pertinent study, researchers discovered that girls' enrollment in high school physics courses was higher in communities where women were employed in nearby STEM occupations (Riegle-Crumb and Moore, 2014). An analogous effect may be occurring with our sample. Of course, this interpretation is speculative and requires testing in future research.

Another potential explanation for parents' greater average science talk with daughters than sons is that parents were responding to subtle gender differences in children's behavior (Bell, 1968). On average, girls tend to do somewhat better in reading (Robinson and Theule Lubienski, 2011) and to be more talkative during early childhood (Leaper and Smith, 2004). Although we did not find average gender differences in children's overall science talk, perhaps girls were more likely receptive to shared book reading and parents found it easier to engage them in science talk. If so, why was parents' science talk more likely among girls (vs. boys) specifically while reading the physical science book? Perhaps engaging the child's interest was more challenging while discussing the more abstract physical science books than the more concrete life science books. Once again, these are speculations that need testing.

In contrast to several prior studies indicating gender-differentiated socialization was more likely among fathers than mothers (refer to Leaper, 2015 for review), we did not observe this in the results. We did observe that fathers were more likely than mothers to make scientific inferences while reading the life science book. Perhaps fathers were generally more comfortable to make these more abstract and cognitively demanding forms of talk (e.g., Tenenbaum and Leaper, 1998). Fathers could also be less familiar than mothers with their children's cognitive abilities, which might lead to using more complex talk. Given that the science inference code was infrequent, these interpretations should be viewed cautiously.

One notable limitation of our study is that our sample was comprised of parents from highly educated backgrounds who also regularly read to their children. Prior study has noted that parents' shared book reading is positively correlated with their education (e.g., Yarosz and Barnett, 2001). Therefore, shared book reading is not a common activity in all families. Another limitation was that the parents in our study were predominantly from White European-heritage backgrounds. Other research suggests that gender-differentiated talk in reading and other learning tasks may vary across different ethnic or cultural groups (e.g., Shirefley et al., 2020).

CONCLUSION

Our findings pose new research questions regarding parents' gender-differentiated encouragement of children's science interest, confidence, and achievement. In contrast to prior studies, we discovered that parents used the more scientific talk with their daughters than sons. This was especially likely with the physical science book. Although more research is needed to replicate and better understand the results, one possibility is that shared book reading could be a learning context conducive for promoting science interest in many girls. To test this premise, short-term longitudinal studies could examine whether this type of book reading in early childhood is related to an increase

in girls' interest in physical science. Moreover, similar benefits may accrue to boys and thereby help all children's developing interest in science.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of California, Santa Cruz Institutional Review Board. Written informed consent to participate in this study was provided by the adult participants and the legal guardian of the minor participants.

AUTHOR CONTRIBUTIONS

TS was largely responsible for data collection, coding, and the data analyses. Both authors worked on writing the report and developed the research questions and methodology.

FUNDING

This research was supported by grants to CL from the Spencer Foundation (grant 201900111) and the UCSC Academic Senate (grant 36146).

ACKNOWLEDGMENTS

The authors are grateful to the parents and the children for their time and participation and also appreciative of Tessa Blood, Kendra Cornell-Rath, Dena Gohari, and Rebecca Upton for their help during data collection, transcription, and coding.

REFERENCES

Anderson, J., Anderson, A., Lynch, J., and Shapiro, J. (2004). Examining the effects of gender and genre on interactions in shared book reading. *Read. Res. Instruct.* 43, 1–20. doi: 10.1080/19388070409558414

Baram-Tsabari, A., and Yarden, A. (2008). Girls' biology, boys' physics: evidence from free-choice science learning settings. *Res. Sci. Technol. Educ.* 26, 75–92. doi: 10.1080/02635140701847538

Bell, R. Q. (1968). A reinterpretation of the direction of effects in studies of socialization. *Psychol. Rev.* 75, 81–95. doi: 10.1037/h0025583

Blake, J., Macdonald, S., Bayrami, L., Agosta, V., and Milian, A. (2006). Book reading styles in dual-parent and single-mother families. *Br. J. Educ. Psychol.* 76, 501–515. doi: 10.1348/000709905X49719

Bronfenbrenner, U. (2005). *Making Human Beings Human: Bioecological Perspectives on Human Development*. Thousand Oaks, CA: Sage.

Bussey, K., and Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychol. Rev.* 106, 676–713. doi: 10.1037/0033-295X.106.4.676

Cheryan, S., Ziegler, S. A., Montoya, A. K., and Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychol. Bull.* 143, 1–35. doi: 10.1037/bul0000052

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn, Mahwah: Lawrence Erlbaum.

Conner, D. B., Knight, D. K., and Cross, D. R. (1997). Mothers' and fathers' scaffolding of their 2-year-olds during problem-solving and literacy interactions. *Br. J. Dev. Psychol.* 15, 323–338. doi: 10.1111/j.2044-835X.1997.tb00524.x

Crowley, K., Callanan, M. A., Tenenbaum, H. R., and Allen, E. (2001). Parents explain more often to boys than to girls during shared scientific thinking. *Psychol. Sci.* 12, 258–261. doi: 10.1111/1467-9280.00347

Eccles, J. S., and Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. *Contemp. Educ. Psychol.* 61:13. doi: 10.1016/j.cedpsych.2020.101859

Endendijk, J. J., Groeneveld, M. G., van der Pol Lotte, D., van Berkel, S. R., Hallers-Haalboom, E., Mesman, J., et al. (2014). Boys don't play with dolls: mothers' and fathers' gender talk during picture book reading. *Parent. Sci. Pract.* 14, 141–161. doi: 10.1080/15295192.2014.972753

Friedman, C. K., Leaper, C., and Bigler, R. S. (2007). Do mothers' gender-related attitudes or comments predict young children's gender beliefs? *Parent. Sci. Pract.* 7, 357–366. doi: 10.1080/15295190701665656

Kelemen, D., Emmons, N. A., Seston Schillaci, R., and Ganea, P. A. (2014). Young children can be taught basic natural selection using a picture-storybook intervention. *Psychol. Sci.* 25, 893–902. doi: 10.1177/0956797613516009

Kenny, D. A., Kashy, D. A., and Cook, W. L. (2006). *Dyadic Data Analysis*. New York, NY: Guilford Press.

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.

Leaper, C. (2015). "Gender and social-cognitive development," in *Handbook of Child Psychology and Developmental Science: Cognitive Processes*, 7th Edn, Vol. 2, eds R. M. Lerner, L. S. Liben, and U. Muller (New York, NY: Wiley), 806–853.

Leaper, C., Anderson, K. J., and Sanders, P. (1998). Moderators of gender effects on parents' talk to their children: a meta-analysis. *Dev. Psychol.* 34, 3–27. doi: 10.1037/0012-1649.34.1.3

Leaper, C., and Smith, T. E. (2004). A meta-analytic review of gender variations in children's language use: talkativeness, affiliative speech, and assertive speech. *Dev. Psychol.* 40, 993–1027. doi: 10.1037/0012-1649.40.6.993

National Science Foundation (2021). *Women, Minorities, and Persons With Disabilities in Science and Engineering*. Washington, DC: National Science Foundation.

Riegle-Crumb, C., and Moore, C. (2014). The gender gap in high school physics: considering the context of local communities. *Soc. Sci. Q.* 95, 253–268. doi: 10.1111/ssqu.12022

Robinson, J. P., and Theule Lubienski, S. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: examining direct cognitive assessments and teacher ratings. *Am. Educ. Res. J.* 48, 268–302. doi: 10.3102/0002831210372249

Scholastic Inc (2016). *Kids and Family Reading Report 2016*. Available online at: http://www.scholastic.com/readingreport/key-findings.htm (accessed October 20, 2021).

Schwartz, J. I. (2004). An observational study of mother/child and father/child interactions in story reading. *J. Res. Childhood Educ.* 19, 10–105. doi: 10.1080/02568540409595058

Shirefley, T. A., Castañeda, C. L., Rodriguez-Gutiérrez, J., Callanan, M. A., and Jipson, J. (2020). Science conversations during family book reading with girls and boys in two cultural communities. *J. Cogn. Dev.* 20, 551–572. doi: 10.1080/15248372.2020.1797750

Short-Meyerson, K., Sandrin, S., and Edwards, C. (2016). Gender influences on parent-child science problem-solving behaviors. *J. Res. Childhood Educ.* 30, 334–348. doi: 10.1080/02568543.2016.1178194

Simpkins, S. D., Fredricks, J., and Eccles, J. S. (2015). The role of parents in the ontogeny of achievement-related motivation and behavioral choices. *Monogr. Soc. Res. Child Dev.* 80, 1–151. doi: 10.1111/mono.12158

Tenenbaum, H. R., and Leaper, C. (1998). Gender effects on Mexican-descent parents' questions and scaffolding during toy play: a sequential analysis. *First Lang.* 18, 129–147. doi: 10.1177/014272379801805301

Tenenbaum, H. R., and Leaper, C. (2003). Parent-child conversations about science: the socialization of gender inequities? *Dev. Psychol.* 39, 34–47. doi: 10.1037/0012-1649.39.1.34

Tenenbaum, H. R., Snow, C. E., Roach, K. A., and Kurland, B. (2005). Talking and reading science: longitudinal data on sex differences in mother-child conversations in low-income families. *J. Appl. Dev. Psychol.* 26, 1–19. doi: 10.1016/j.appdev.2004.10.004

Yarosz, D. J., and Barnett, W. S. (2001). Who reads to young children?: Identifying predictors of family reading activities. *Read. Psychol.* 22, 67–81. doi: 10.1080/02702710151130235

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 February 2022
doi: 10.3389/fpsyg.2022.829333

Gender-Specific Aspects of Teachers Regarding Working Behavior and Early Retirement

Steffi Kreuzfeld[*†] *and Reingard Seibt*[†]

Institute for Preventive Medicine, Rostock University Medical Center, University of Rostock, Rostock, Germany

Worldwide, a significant proportion of teachers retires prematurely for health reasons or at their own request. The study examines whether male and female teachers differ in terms of working conditions and coping with high work demands as well as individual factors that promote early retirement. A cross-sectional study was conducted to collect data from 6,109 full-time teachers in high schools (56% women). Weekly working hours from a four-week working time record and psychosocial work stress (effort-reward model, ER ratio) were used as workloads. In addition, emotional exhaustion (Maslach Burnout Inventory) and coping strategies that endangered health were recorded in the form of overcommitment and inability to recover. Also, the teachers gave a prediction and reasons for early retirement and made their own suggestions on how to prevent this. The results show that both workloads and emotional exhaustion are comparable between the genders, but women have a greater tendency than men to overcommit and be unable to recover. As ER ratio and emotional exhaustion increase, the chances for both genders to reach the regular retirement age decrease significantly; for health-endangering coping strategies, the relationship is somewhat weaker. The majority of male and female teachers (79%) indicates excessive workloads as the main reason for leaving the profession early. In order to protect teachers from high workloads, measures at the organizational, social, and individual level are necessary. Proposals for schools and policy makers are critically discussed on the basis of teacher recommendations.

Keywords: teachers, gender, overcommitment, recovery, retirement

OPEN ACCESS

Edited by:
Daniela Barni,
University of Bergamo, Italy

Reviewed by:
Ilaria Buonomo,
Libera Università Maria SS. Assunta, Italy
Johannes Wendsche,
Federal Institute for Occupational Safety and Health, Germany

***Correspondence:**
Steffi Kreuzfeld
steffi.kreuzfeld@uni-rostock.de

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *05 December 2021*
Accepted: *06 January 2022*
Published: *15 February 2022*

Citation:
Kreuzfeld S and Seibt R (2022) Gender-Specific Aspects of Teachers Regarding Working Behavior and Early Retirement. Front. Psychol. 13:829333. doi: 10.3389/fpsyg.2022.829333

INTRODUCTION

The teaching profession is characterized by a complex structure of work demands and stressors. In addition to high mental, emotional, and psychosocial work load (Shirom et al., 2009; Skaalvik and Skaalvik, 2017a; Framke et al., 2021), the profession can be described by a high degree of autonomy. Compared to other occupational groups, there is an increased risk of stress-related psychosomatic and mental illnesses, including burnout (Guglielmi and Tatrow, 1998; García-Carmona et al., 2019). On the other hand, teachers are often more satisfied with their jobs than other professional groups despite the high demands (Schult et al., 2014).

With regard to work organization, the activities of the teachers show typical characteristics of the flexible working world with individualized working hours and locations. This goes hand

in hand with a high level of personal responsibility for the results of the work, the risk of permeable boundaries between job and private life (Ashforth et al., 2000; Clark, 2000) and the risk of time and performance pressure hazardous to health (Höge and Hornung, 2013). However, many teachers appreciate the high degree of autonomy in their profession. In addition to a fixed number of teaching hours and regular extracurricular appointments (e.g., meetings), they are relatively free to allocate more than half of their working time (European Commission/Eurydice, 2013; OECD, 2019). However, the prerequisite for successfully coping with the extremely diverse work tasks is professional self-organization through which professional time and expenditure are individually controlled. In this respect, the subjective potential and resources of a teacher (e.g., work organization and ability to recover) are of considerable importance for the long-term preservation of health and work ability.

A conflict between self-determination and self-endangerment results from weighing up of one's own quality standards at work and the need to recover from work. High expectations placed on teachers by society, parents, and students and their own desires for success, e.g., good student performance, can cause excessive exertion even though they realize that they are putting their health at risk (Dettmers et al., 2016). Krause et al. (2015) coined the term "interested self-endangerment" to describe this behavior. Typical examples of self-endangering behavior include presentism, working excessively long hours, working on weekends and vacations, and not taking recovery breaks. In a recent article, the working group around Krause demonstrated that teachers achieve short-term success in coping with their work demands by extending their working hours and thus experience themselves as competent. However, in the long term, they increase the risk of mental health impairments if they ignore the recovery required (Baeriswyl et al., 2021).

One model that analyzes the impact of job demands on teacher health is the effort-reward (ER) model (Siegrist et al., 2009). High effort is caused by working under time pressure, working with interruptions, or by an increase in workload. Reward, on the contrary, subsumes both material aspects, such as salary, promotion opportunities, and job security, and immaterial aspects, such as appreciation by colleagues, superiors, students, and parents. In a favorable case, for example, teachers' high workload is compensated by adequate pay and appreciation by colleagues and students. According to this model, an effort-reward imbalance (ERI) generates psychosocial work stress that increases the risk of stress-associated illness and burnout in the medium and long term (Van Vegchel et al., 2005, Lehr et al., 2009; Zurlo et al., 2010; Wang et al., 2015; Rugulies et al., 2017; Solis-Soto et al., 2019; Madsen and Rugulies, 2021). In the study by Loerbroks et al. (2014), the ERI score in elementary school teachers was found to be a strong determinant not only of burnout but also of intention to leave the profession. Niedhammer et al. (2014) postulate that 16% of mental disorders can be attributed to an ERI.

In addition to work stress (extrinsic factor), the ER model also includes overcommitment (OC) as an intrinsic component. This describes an individual coping style with the tendency to overexert oneself without regard to one's own resources (Siegrist and Li, 2016). It assumes that teachers who are simultaneously characterized by high ERI and high OC are at the greatest risk of decreased health and wellbeing (Siegrist et al., 2009; Siegrist and Li, 2016; Hinsch et al., 2019). However, the evidence on this is inconsistent. Excessive work engagement can, on the one hand, lead to extended working hours and, on the other hand, have unfavorable effects on recovery processes during non-working time, e.g., by not fully compensating for the consequences of previous activities (Meijman and Mulder, 1998; Sonnentag, 2003). Both long working hours and shortened recovery times are relevant to health (Meijman and Mulder, 1998; Wepfer et al., 2018).

Teachers working full-time in Germany often have high weekly working hours (Ø 45 h/week; Felsing et al., 2019a). Many even have to regularly use the weekend to manage their workload. In doing so, longer recovery intervals are fragmented. As a result, teachers have more limited recovery opportunities than people in other occupations with shorter working weeks or part-time work.

Recovery processes may also be impaired in teachers if they are unable to sufficiently distance themselves from work-related content (Sonnentag and Fritz, 2015). This mental detachment from work during rest periods is seen as a central component of individual recovery (Sonnentag and Fritz, 2015; Stieler et al., 2019). It is considered a link between working conditions and stress-related outcomes and has been discussed as an early indicator of exhaustion and burnout (Wendsche and Lohmann-Haislah, 2017; Seibt and Kreuzfeld, 2021). A physiological activation which lasts for the duration of working time is seen as a pathomechanism (Sonnentag and Fritz, 2015) which hinders the necessary recovery associated with persistent cognitive processes, such as affective rumination (Rau and Triemer, 2004; Cropley and Zijlstra, 2011). Also, excessive work engagement can thus contribute to individuals losing the ability to relax. This inability to recover is considered an individual pattern in coping with work demands and is considered a health risk factor in its own right among teachers (Varol et al., 2021). In a longitudinal study, Sonnentag et al. (2010) were able to predict future exhaustion for employees who have poor mental detachment in their free time. Exhaustion in this respect is the result of a chronic overtaxing of one's own performance reserves. In the "burnout concept" according to Maslach and Jackson (1981), emotional exhaustion is seen as the core component (Hakanen et al., 2006; Skaalvik and Skaalvik, 2011a).

Gender seems to play an important role in relation to the extent of work stress. However, previous studies have provided contradictory findings in this regard (Gyllensten and Palmer, 2005). Women and men both differ in the way they are exposed to stress and in their response to stress (Folkman et al., 1986; Arntén et al., 2008). Causal factors include differences in working conditions, social role behavior, role conflicts, especially work-family conflicts, gender stereotypes, and related differences in advancement opportunities, among others (Gyllensten and Palmer, 2005; Li et al., 2006). In principle, workplace stress and work-family conflicts are risk factors for mental health disorders in both genders (Wang et al., 2008). For women, however, the probability of work-family conflicts and emotional

exhaustion increases when they have to work longer than desired, i.e., they suffer from over-employment (Rubino et al., 2013).

Research has considered the differential effects of stress on men's and women's health under two hypotheses: differential exposure and differential vulnerability. The first hypothesis assumes that with fewer work-related resources (e.g., income and job promotion), women are exposed to interpersonal, emotional, and social stressors, including work-family conflicts, to a greater extent than men and therefore complain more about stress-related problems (e.g., Bond et al., 2004). The second hypothesis assumes that women react more sensitively to certain stressors because of the additive effect of family and paid work roles (Roxburgh, 1996; Arroba and James, 2002; Liu et al., 2008). It is conceivable that the observed gender differences will be reduced by the convergence of social roles (e.g., fathers taking on more family responsibilities) and the equalization of working conditions (Frankenhaeuser, 1991; Persson et al., 2008).

Decades ago, there were some studies that found gender differences in stress triggers and perceptions of stress levels among teachers as well (Laughlin, 1984; Travers and Cooper, 1991). In contrast, other studies found no gender-based differences (Jepson and Forrest, 2006; Reilly et al., 2014). Subsequently, the most common causes of occupational stress among teachers were identified as high work demands, student misbehavior, lack of student interest and motivation, and difficult interactions with colleagues and parents (Borg, 1990; Klassen et al., 2012; Aldrup et al., 2018). Here, female teachers have reported significantly higher levels of occupational stress than their male colleagues particularly in interactions with students and colleagues (Griffith et al., 1999; Antoniou et al., 2006). Females also reported higher levels of workload and emotional exhaustion compared to their male counterparts (Van Dick and Wagner, 2001; Sünbül, 2003; Antoniou et al., 2006; Wang et al., 2015; Arvidsson et al., 2016), and more discomfort and a higher anxiety level (Tamres et al., 2002; Chong and Chan, 2010; Arvidsson et al., 2016). Overall, female teachers perceive stress more often (Greenglass and Burke, 2003; Rasku and Kinnunen, 2003; Antoniou et al., 2006; Chaplain, 2008; Agaiâ-Demjaha et al., 2015) and rate their health worse than their male colleagues (Lagrosen and Lagrosen, 2020). In addition to gender, the number of working hours and relationship status are further factors influencing the extent of emotional exhaustion among teachers. Thus, married and partnered teachers as well as teachers with a weekly working time of less than 40 h each reported a lower emotional exhaustion than singles and employees with a weekly working time of more than 40 h (Wang et al., 2015).

Some teachers only become aware of the finite nature of their own resources when they are emotionally exhausted or suffer burnout. Latest now, there is a real risk of them having to give up the teaching profession and taking early retirement. Mental illnesses, especially emotional exhaustion, are closely related to early retirement among teachers (Leung and Lee, 2006; Skaalvik and Skaalvik, 2011b). Early retirement is understood here as the time of complete withdrawal from the teaching profession before reaching the official retirement age. In research, retirement is viewed as a process that examines retirement planning and the decision to retire, as well as retirement with its corresponding consequences (Beehr, 1986; Topa et al., 2009; Fisher et al., 2016a; Topa et al., 2018). This very complex process is influenced by a variety of individual, family, and work-related factors. In this context, personal goals interact with financial and health constraints.

In Germany, the regular retirement age for teachers is 67 years. However, only about one in four teachers reaches the statutory retirement age (Statistisches Bundesamt, 2018). Concrete data on the proportion of teachers who retire early due to invalidity are not available for Germany. A significant proportion of teachers still leave the profession early at their own request. According to research by Van Droogenbroeck and Spruyt (2014), female teachers are more likely than male teachers to want to retire.

In summary, it can be said that working conditions as well as individual and health-related factors can promote early retirement among teachers. With regard to previous studies on work-related stress among teachers, it should be critically noted that working conditions were often the focus of studies and that very heterogeneous samples are examined, including both full-time and part-time employees as well as teachers from primary and high schools, and teachers with different job profiles (e.g., principals, teachers with special functions, and regular teachers). This severely limits the interpretability of the results and may lead to incorrect conclusions. Therefore, further studies with homogeneous teacher samples are needed to analyze the causes of gender differences in work-related stress outcomes, paying particular attention to individual factors.

The aim of the study was therefore to identify possible gender differences in work, personal, and health characteristics on the basis of a dataset that is representative for full-time high school teachers in Germany. In particular, aspects of self-harming behavior were to be considered. Furthermore, it was necessary to clarify the question of the predictive value individual characteristics have for the probability of reaching regular retirement age. In addition, the analysis covered whether the subjective reasons for early retirement differed between female and male teachers.

MATERIALS AND METHODS

Procedures and Data Collection

The data for the present study were collected as part of the Germany-wide, cross-sectional study "Lehrerarbeit im Wandel" (Teaching under Change - LaiW study) between January and April 2018. The study determined the workload and health of high school (in German: Gymnasium) teachers in all 16 German states. The study period selected for the individual federal state represented an average workload in each case (no extraordinary activities, such as exams or extensive correction work).

In the run-up to the study, posters and flyers were placed at all high schools to advertise voluntary participation. Before the start of the study, all teachers received an information letter on the study in their school with information on data

protection, implementation, and data evaluation as well as on the conditions for participation and access to the study. Anonymity of the data was ensured *via* transaction numbers and an eight-digit personal code. Data collection took place *via* an online portal of the University of Rostock.

A list of answers to frequently asked questions was available to participants on the LaiW study website for queries. In addition, the study team could be contacted by telephone and electronically throughout the study period.

The study consisted of an online questionnaire (OQ) and an online protocol (OP). First, all participants answered the online questionnaire on sociodemographic, job-specific, and health-related questions once. Subsequently, they logged their working time daily in the online protocol over a period of 4 weeks (28 days) using defined activity categories. From this, an average weekly working time was determined.

More than 20,000 high school teachers (hereafter teachers) participated in the LaiW study. Using the personal code, the online questionnaire and online protocol could be merged for data analysis. Complete datasets were available for 14,338 participants due to matching codes. About 84% of these records ($n = 12{,}014$) were related to teachers who primarily give lessons. In contrast, 16% of the records referred to teachers who were employed as head teachers or deputy head teachers or who performed other administrative tasks and functions within the school to a considerable extent and therefore gave significantly fewer lessons. For the comparison of gender-specific aspects, a sample should be studied that was as homogeneous as possible. Therefore, only datasets from full-time teachers with a reduction of up to 3 h (reduced teaching hours) were analyzed ($n = 6{,}109$).

Sample

The sample of 6,109 full-time teachers was composed of a slightly higher proportion of women (56%) compared to men (44%). The mean age of men was 42 ± 10 years and that of women was 41 ± 10 years ($d = 0.180$). The further composition of the sample is summarized in **Table 1**. Please see data analyses for information on the interpretation of effect sizes (d).

Combinations of languages and social sciences as well as languages and natural sciences were taught most frequently, with women indicating more pure language subjects and men indicating more natural science subjects ($d = 0.529$, medium effect).

Most teachers lived in a stable partnership (men: 87%, women: 77%; $d = 0.252$, small effect). About 5% of teachers reported having to care for relatives in the household, and more than a third (35%) of them also took care of children in their own household; however, this applied to 52% of men but only 21% of women ($d = 0.719$, medium effect).

Measures

Working time and the characteristics of psychosocial workload are used to describe workload (Siegrist et al., 2009). Overcommitment (Siegrist et al., 2009) and inability to recover (Richter et al., 1996) are attributes of self-harming behavior and emotional exhaustion is considered a health-related characteristic (Schaufeli et al., 1996). All questions about future retirement were developed in-house.

Online Questionnaire

In addition to sociodemographic (e.g., gender, age, and marital status) and occupation-specific information about teachers (e.g., teaching responsibilities, subjects taught, classes, and number of students), the OQ also included questions about work, personal, and health characteristics. Standardized questionnaires and supplementary self-developed questions were used to record these characteristics.

Psychosocial Workload

It was surveyed with the Effort-Reward-Imbalance Questionnaire (ERI-Q: Siegrist et al., 2009). This questionnaire allows the standardized measurement of occupational gratification crises. The short version used by Siegrist et al. (2009) included the main scales effort (3 items; range: 3–15 points) and reward (7 items; range: 7–35 points), as well as the effort-reward ratio (ER ratio). Each effort item was measured on a five-point scale from 1 ("disagree") to 5 ("agree, and I'm very distressed"). The reward scale consisted of three subscales: status or job promotion, valuation or esteem, and job security. The reward items were measured on a five-point scale from 1 ("agree) to 5 ("disagree, and I'm very distressed"). High sum values indicated high perceived effort or reward. The ER ratio was formed from the sum values of the two main subscales by the following rule: ER-Ratio = $\sum$ effort/($\sum$ reward * 0.54). An ER ratio of >1 indicates an ERI (Siegrist et al., 2009), which is said to be associated with a health risk. The greater the imbalance between effort and reward (gratification crisis), the higher the health risk is said to be. Validity and reliability of the German short version of ERI-Q were satisfactory (Siegrist et al., 2009). For the main subscales, the values of internal consistency were above 0.70 (effort: 0.74 and reward: 0.79). For the ER scales of the present study, lower Cronbach Alpha's were determined (effort: 0.61 and reward: 0.72), which can be classified as questionable or acceptable (Blanz, 2015).

Overcommitment

Overcommitment (OC) or excessive work commitment was also assessed with the short version of the ERI-Q (Siegrist et al., 2009). This is an individual coping style with a tendency to spend oneself without regard to one's resources. The OC scale comprises six items that are rated on a four-point Likert scale (1 = strongly disagree up to 4 = strongly agree). In this scale, a sum score is formed from the six items (value range: 6–24 points), in which high values correspond to a high propensity to exert oneself. The upper tercile of the sum score was defined as the risk group (Siegrist et al., 2004). A Cronbach's Alpha of 0.79 is given as the internal consistency of the OC subscale (Siegrist et al., 2009). In the present LaiW study, the Cronbach's Alpha for OC was 0.77, which is acceptable (Blanz, 2015).

Inability to Recover

Inability to recover (IR) is a subscale of Questionnaire for Faulty Attitudes and Behavior Analysis relevant to coping with work demands (Richter et al., 1996). Depicted is extreme work

TABLE 1 | Characteristics of the sample of male and female full-time teachers.

	Full-time teacher				Significance	
	Male (n = 2,680)		Female (n = 3,429)		Test value	Value of p (effect size)
	%	n	%	n		
Age groups [years]	5.1	138	14.5	498		
25–29	39.9	1,069	40.0	1,373	215.51	<0.001 (0.382)
30–39	31.3	840	20.3	695		
40–49	17.5	469	20.6	708		
50–59	6.1	164	4.5	155		
60–67	5.1	138	14.5	498		
Subjects and subject combinations						
Languages	7.9	213	22.7	777	398.95	<0.001 (0.529)
Social sciences	5.6	149	2.1	71		
Natural sciences	27.5	736	16.7	574		
Languages and social sciences	23.8	638	25.1	859		
Languages and natural sciences	3.7	99	8.0	274		
Social sciences and natural sciences	8.6	230	7.0	241		
Art, music, sports	2.5	66	1.7	60		
Subject combinations with art, music, sports	20.5	549	16.7	573		
Family obligations						
Permanent partnership	86.7	2,323	76.8	2,635	95.15	<0.001 (0.252)
Children in the household	52.0	1,394	21.4	733	699.16	<0.001 (0.719)
Care of relatives	4.5	120	5.8	198	5.13	0.024 (0.058)

%: frequencies in %; n: number of teachers. Chi-square test according to Pearson (test size: χ^2-value, effect size: d); p-value: significance (two-sided). Effect size according to Cohen (1988): d: <0.20 = no effect, 0.20–0.49 = small effect, 0.50–0.79 = medium effect, and ≥0.80 = large effect.

commitment associated with accepted limited recovery ability in terms of an inefficient coping style (Richter et al., 1999). The inability to recover is assessed with six items using a four-point ranking scale (1 = not at all true to 4 = very true). Then, the sum value (range: 6–24 points) is formed over the six items, which can be assigned on the basis of percentile values to normal (6–18 points), high (19–21 points), and very high (22–24 points) recovery values. The reliability of the IR subscale was reported by Richter et al. (2015) with a Cronbach's Alpha of 0.79. The Cronbach's Alpha for IR was calculated to be 0.82 in our study and can be assigned to the good range (Blanz, 2015).

Emotional Exhaustion

Emotional exhaustion (EE) is considered the core component of the frequently cited burnout definition of Maslach and Jackson (1981) and was recorded by the German translation of the Maslach Burnout Inventory - General Survey (MBI-GS: Schaufeli et al., 1996). The subscale EE consists of five statements (items), which are assessed on a seven-point Likert scale (0 = never up to 6 = daily) according to their frequency of occurrence and are summarized as a mean value to form the EE score. High EE scores indicate typical stress reactions and the draining of emotional resources. For evaluation, the mean values of the subscale EE can be classified as low (<2.0 points), average (2.0–3.2 points), and high (>3.2 points; Maslach and Jackson, 1986). The validity evidence of the MBI has been demonstrated both for normal and clinical populations (Schaufeli et al., 2001) and for different occupational groups. Maslach and Jackson (1986) presented internal consistencies in the form of Cronbach's Alphas of 0.90 for emotional exhaustion for a sample of 1,316 subjects. Schaufeli et al. (1996) report a Cronbach's Alpha of 0.78 for emotional exhaustion. For the study presented here, the Cronbach's Alpha was 0.78, which is in the acceptable range according to Blanz (2015).

Time of Retirement

The questions about the probability of early versus regular retirement and about the individual reasons were developed in-house for pragmatic considerations and were each recorded with a global question. The following question was to be answered as: "Can you imagine practicing your profession until the statutory regular retirement age?" If the question was answered "no," a maximum of three main reasons for early retirement was to be given. Similarly, participants were asked to suggest two to three specific actions that they believe were necessary to remain healthy and employed in the teaching profession until regular retirement age. These statements and the reasons for early retirement were free text statements that were manually evaluated or categorized for all 6,109 teachers.

Online Protocol

The OP served to determine the weekly working time and activity structure of the teachers. To do this, the teachers had to document their work time daily for 4 weeks (28 days) using 12 practicable, suitably clear categories of teacher-specific activity, which were grouped into the following higher-level domains:

- teaching (lessons, substitution lessons)
- teaching-related activities (preparation and follow-up of lessons, correction and grading of students' work, marking, preparation of projects, and excursions)
- non-teaching activities (work with students and parents, administration, work with colleagues, tasks within the scope of students' inclusion and integration, supervision time, and all other tasks).

The total weekly working time was calculated by first determining the average values over 4 weeks for each activity category and subsequently summarizing these as the weekly working time. The amount of time for the individual activity categories was previously examined for statistical outliers. Extreme values were replaced with subject-specific mean values within each activity category. Participants who recorded their working time on fewer than 21 of 28 days were not included in the data analysis.

Data Analyses

Prior to the statistical calculations, the entire dataset was checked for implausible data. Input aids and default settings in both the online questionnaire and the online working time log prevented implausible data from being entered.

Statistical analysis of the data was performed with the Statistical Package for the Social Science (SPSS INC, Chicago, IL, United States) for Windows (version 27). A probability of error of $\alpha < 0.05$ was set as the statistical significance criterion and supplemented by effect sizes. The interpretation of effect sizes was based on the conventions of Cohen (1988). Statistically significant effects in the analyses of variance or the χ^2 tests were considered to be small effect sizes from $\eta^2_{partial} = 0.01$ or $d = 0.20$, respectively.

The focus of this paper is on the analysis of gender effects for the examined work-, person-, and health-related characteristics. Mean differences between male and female teachers were examined for these characteristics - after checking for age groups and subject profiles - using univariate General Linear Models. The χ^2 test was used for difference testing of categorical variables.

Correlations between work-related and personal characteristics and emotional exhaustion with the variable early or regular retirement were examined gender-specifically and with point-biserial correlations. Correlations between the characteristics were analyzed using Pearson product-moment correlation. Correlation coefficients were interpreted according to Bühl (2016), where $r \pm \leq 0.10$ was considered independent of each other.

Binary logistic regression analyses were carried out to clarify the question of the predictive value work and individual characteristics (independent variables), including control variables, have for the probability of reaching the regular retirement age (response variables). These analyses were performed separately for male and female teachers. The selection of characteristics included in the overall model (method: enter) was based on the results of the correlation analysis; this was prefixed to the regression. To assess the goodness-of-fit, the Nagelkerke R^2 was used, which can assume values of between 0 and 1.

RESULTS

Gender Comparison for Workload

Weekly teaching hours, time for teaching-related and non-teaching activities, and working time were investigated as working time-related characteristics (see **Table 2**). As expected, the number of compulsory hours does not differ between male and female teachers ($p = 0.234$); they teach an average of 22 school hours per week (á 45 min). For teaching-related activities, however, women report an average of 19 h/week, about 2 h more than men ($\eta^2_{partial} = 0.016$, small effect), while there is only a marginal gender effect for the time spent on non-teaching activities ($\eta^2_{partial} = 0.003$); on average, all teachers invest 10 h/week for these tasks. In summary, female teachers work an average of 1.5 h more per week than male teachers (∅ 45.7 vs. 44.2 h/week; $\eta^2_{partial} = 0.012$, small effect).

For the control variables age group and subject profile, there are statistically significant effects ($p < 0.05$) for activity proportions and working time, but it is not practically significant ($\eta^2_{partial} < 0.010$). Younger colleagues (20–29 years) nevertheless have significantly longer working hours than older colleagues (60–67 years; ∅ 47.0 vs. 42.2 h/week; $\eta^2_{partial} = 0.012$, small effect). When looking at it in terms of gender, this difference can only be confirmed for female teachers (∅ 47.4 vs. 42.9 h/week; $\eta^2_{partial} = 0.011$, small effect).

The effort-reward subscales (ER subscales) are considered a second aspect of workload (see **Table 2**). The mean scores of the ER subscales are not significantly different between male and female teachers ($p \geq 0.05$). The mean scores of all teachers are still in the normal range for effort (∅ 10 of 15 points) and reward (∅ 26 of 35 points). For the ER ratio, the mean value of the teachers is 0.93 and thus still outside the risk range. Nevertheless, there is a health risk due to the imbalance of effort and reward (ER ratio > 1) for more than one-third of them (35%). The three ER subscales effort, reward, and ER ratio are not influenced by the subject profiles taught ($\eta^2_{partial} < 0.010$). However, there is an age effect for the results on effort ($\eta^2_{partial} = 0.019$, small effect).

Neither gender differences nor age effects nor effects related to the subject profile can be determined with significant practical importance for the three reward subscales promotion, esteem, and job security ($\eta^2_{partial} < 0.010$). The opportunities for job promotion are reported by the teachers with an average of 11 points (range: 3–15 points). Perceived professional esteem is rated an average of 7, and job security with 8 points (range in each case: 2–10 points).

Gender Comparison for Personal Characteristics

Overcommitment and inability to recover were investigated as person-related characteristics with a link to self-harming behavior. On average, the mean scores for both overcommitment and inability to recover differ between male and female teachers (see **Table 3**; $\eta^2_{partial} \geq 0.01$, small effects). These are within the normal range for both genders and both characteristics but are close to the border of the high range (>18 of 24 points).

TABLE 2 | Main effects of work-related characteristics and covariates (age groups and subject profile) for male and female full-time teachers.

	Full-time teacher			Significance		
	Dimension	**Male (n = 2,680)**	**Female (n = 3,429)**	***F*-value**	**Value of *p***	**Effect sizes ($\eta^2_{partial}$, d)**
Workload						
Teaching [hours/week, á 45 min]	M ± SD	22.5 ± 3.5	22.7 ± 3.5	1.42	0.234	0.001
Age group				62.16	<0.001***	0.009
Subject profile				0.17	0.680	0.001
Teaching-related activities [hours/week]	M ± SD	17.3 ± 6.6	19.4 ± 7.1	96.68	<0.001***	0.016
Age group				24.64	<0.001***	0.004
Subject profile				140.43	<0.001***	0.022
Non-teaching activities [hours/week]	M ± SD	10.0 ± 3.5	9.3 ± 3.1	20.94	<0.001***	0.003
Age group				57.69	<0.001***	0.009
Subject profile				7.93	<0.005**	0.001
Working time [hours/week]	M ± SD	44.2 ± 8.6	45.7 ± 8.7	31.93	<0.001***	0.005
Age group				43.94	<0.001***	0.007
Subject profile				13.22	<0.001***	0.002
Effort-reward subscales						
Effort [5–15 pts]	M ± SD	9.7 ± 2.6	9.5 ± 2.6	0.89	0.345	0.001
Age group				121.7	<0.001***	0.019
Subject profile				1.16	0.345	0.001
Reward [7–35 pts]	M ± SD	26.0 ± 5.4	26.0 ± 5.4	0.00	0.955	<0.001
Age group				20.45	<0.001***	0.003
Subject profile				8.22	0.004**	0.004
Effort-reward ratio (ER ratio)	M ± SD	0.93 ± 0.42	0.92 ± 0.43	0.19	0.665	<0.001
Age group				62.92	<0.001***	0.009
Subject profile				7.12	0.008**	0.001
Evaluation of ER ratio						
ER ratio ≤ 1	% (n)	64.3 (1,724)	66.1 (2,266)	2.04	0.153	0.037
ER ratio > 1	% (n)	35.7 (956)	33.9 (1,163)			

*pts: points; M ± SD: mean ± standard deviation; % (n): frequency in %, n: number of teachers. Chi-square test according to Pearson (test size: χ^2-value, effect size: d); univariate analyses of variance, design: constant term, sex + age group + subject profile (test size: F-value, effect size: $\eta^2_{partial}$: partial eta-square); value of p: significance (two-sided): ***p < 0.001, **p < 0.01. Effect size according to Cohen (1988): $\eta^2_{partial}$: <0.01 = no effect, d: <0.20 = no effect. There are no age and subject profile effects for the effort-reward subscales ($\eta^2_{partial}$ <0.010). Corrected R-squared: teaching = 0.009, teaching-related activities = 0.044, non-teaching activities = 0.015, working time = 0.015, effort = 0.015, reward = 0.005, and effort-reward ratio = 0.010.*

TABLE 3 | Main effects of personal characteristics and covariates (age groups and subject profile) of male and female full-time teachers.

Personal characteristics	**Dimension**	**Male (n = 2,680)**	**Female (n = 3,429)**	***F*-value**	**Value of *p***	**Effect sizes ($\eta^2_{partial}$)**
Overcommitment (OC) [6–24 pts]	M ± SD	16.8 ± 3.5	18.1 ± 3.3	207.51	<0.001***	0.033
Age group				4.14	0.042 *	0.001
Subject profile				10.26	0.001***	0.002
Inability to recover (IR) [6–24 pts]	M ± SD	16.5 ± 3.6	17.7 ± 3.4	186.32	<0.001***	0.030
Age group				3.73	0.053	0.001
Subject profile				9.11	0.003**	0.001

*pts: points; M ± SD: mean ± standard deviation; univariate analyses of variance, design: constant term + sex + age group + subject profile; test size: F-value, value of p: significance (two-sided): ***p < 0.001, **p < 0.01, and *p < 0.05. Effect size according to Cohen (1988): $\eta^2_{partial}$: <0.01 = no effect, 0.01–0.05 = small effect. Corrected R-squared: OC = 0.037, IR = 0.031.*

One-third of male teachers (33%) and about half of female teachers (47%) tend to overexert themselves (see **Figure 1**, $d = 0.288$, small effect). A similar pattern emerges for inability to recover. Here too, significantly more female than male teachers show high or very high values ($d = 0.250$, small effect) at 47 and 38%, respectively. Age and subject profile have no relevant influence on these results ($\eta^2_{partial} < 0.010$).

From the perspective of personal characteristics, there is a clear health risk from self-harming behavior for a total of 21% of male teachers and for more than one-third (35%) of female teachers; they are noticable for both high values for overcommitment and inability to recover. Only half of the teachers (51%) show normal levels of overcommitment and recovery at the same time (men: 59% and women: 44%).

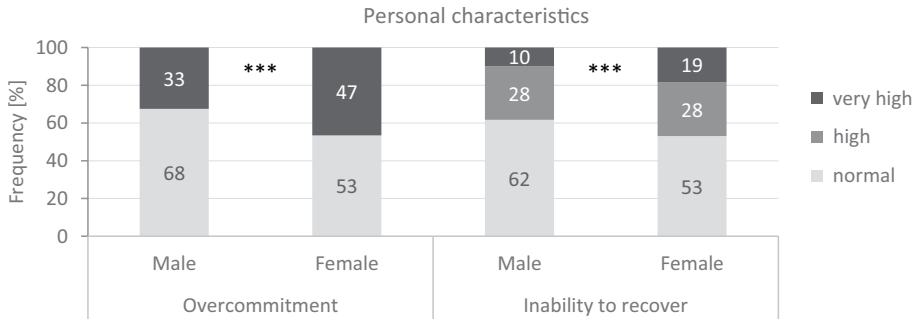


FIGURE 1 | Overcommitment and inability to recover of male ($n = 2{,}680$) and female ($n = 3{,}429$) full-time teachers. Chi-square test according to Pearson (test size: χ^2-value, effect size: d); significance (two-sided): ***$p < 0.001$. Effect size according to Cohen (1988): *d*: 0.20–0.49 = small effect.

Gender Comparison for Emotional Exhaustion

There is a significant difference between male and female teachers ($p < 0.001$) for emotional exhaustion - checking for age groups and subject profile - but this difference is also not practically relevant ($\eta^2_{\text{partial}} < 0.01$; see **Table 4**). Thus, the average values of teachers (range: 0–6 points) for emotional exhaustion are at 2.4 points. According to this, teachers experience emotional exhaustion on average "once a month." This result is not influenced by age effects or effects of the subject profiles taught ($\eta^2_{\text{partial}} < 0.01$).

According to the classification recommended by Maslach and Jackson (1986), the mean values of emotional exhaustion are in the average range for both genders. Just under a quarter (23%) of the male teachers and a third (31%) of the female teachers show high emotional exhaustion (see **Figure 2**).

TABLE 4 | Main effects of emotional exhaustion and covariates (age groups and subject profile) of male and female full-time teachers.

	Full-time teacher		Significance		
	Male ($n = 2{,}680$)	Female ($n = 3{,}429$)	*F*-value	Value of *p*	Effect sizes (η^2_{partial})
Corrected model			20.07	<0.001***	0.010
Constant term			2400.26	<0.001***	0.282
Emotional exhaustion	2.2 ± 1.3	2.5 ± 1.2	59.50	<0.001***	0.009
Age group			0.07	0.796	<0.001
Subject profile			0.07	0.788	<0.001

*pts: points; means ± standard deviations; univariate analysis of variance, design: constant term, sex + age group + subject profile (test size: F-value (Fishers F), effect size: η^2_{partial}: partial eta-square); value of p: significance (two-sided): ***$p < 0.001$. Effect size according to Cohen (1988): η^2_{partial}: <0.01 = no effect, 0.01–0.05 = small effect, and ≥0.14 = large effect. Corrected R-squared = 0.009.*

Gender Comparison for the Time of Retirement

The question about early retirement was answered significantly differently by male and female teachers ($d = 0.261$, small effect): 30% of male and 42% of female teachers estimate that they will not remain in the profession until they reach the regular retirement age. The reasons for this hardly differ between the two gender groups ($d < 0.20$; see **Figure 3**). About half (51%) of the teachers made two statements and 13% made three. More than three quarters (79%) cited excessive workload as the main reason for taking early retirement. Age-related decrease in physical strength (18%) or mental illness (including emotional exhaustion; 18%) was also cited. For a few female teachers (<1%), caring for relatives is also a reason for early retirement. About 3% of teachers provided incorrect or no information on reasons for early retirement. Further reasons are shown in **Figure 3**.

In order to reach the regular retirement age in good health, teachers suggest the following main measures: reducing the number of compulsory hours (46%), decreasing class size (29%), reducing additional tasks (21%) as well as bureaucracy and the administrative burden (18%), and improving organizational conditions (22%). No relevant gender differences could be demonstrated for any of the proposed measures ($d < 0.20$).

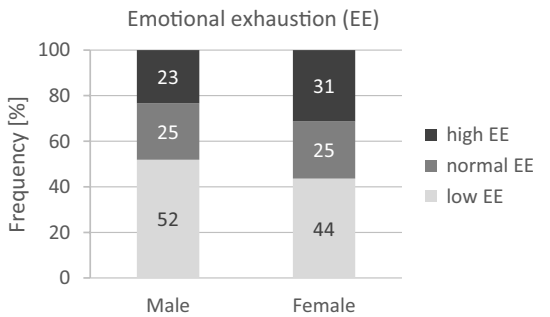


FIGURE 2 | Emotional exhaustion of male ($n = 2{,}680$) and female ($n = 3{,}429$) full-time teachers. Chi-square test according to Pearson (test size: χ^2-value, effect size: d); significance (two-sided). Effect size according to Cohen (1988): *d*: <0.20 = no effect.

Associations Between Work-, Person-, and Health-Related Characteristics With Retirement

The correlation analyses examined the relationship between work-, person-, and health-related characteristics and age with the variable early versus regular retirement for both genders.

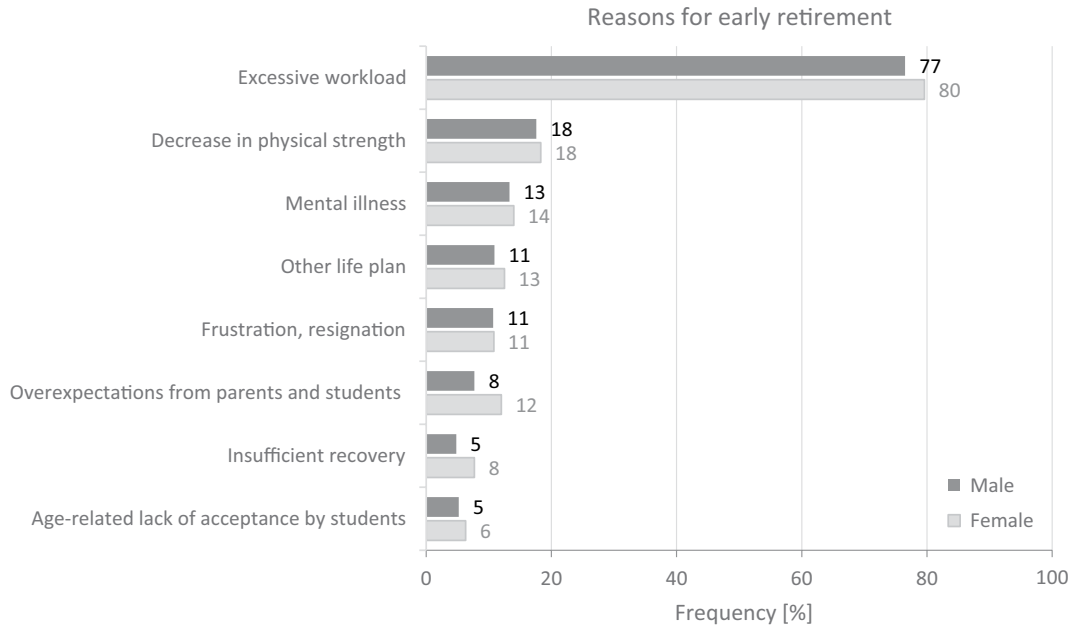


FIGURE 3 | Reasons for not reaching regular retirement of male ($n = 2{,}680$) and female ($n = 3{,}429$) full-time teachers (multiple responses possible).

The strength of the examined correlations did not differ between the genders. The work-related characteristics do not show any statistical significant correlation with the variable retirement start date ($r = -0.08$–0.01). The effort-reward subscales correlate low with prediction of retirement ($r = -0.31$ to -0.28), i.e., low occupational effort, high reward, and low ER ratio tend to be associated with attainment of regular retirement age.

The same applies to the correlations between retirement and personal characteristics ($r = -0.29$ to -0.21) or emotional exhaustion (male: $r = -0.35$, female: $r = -0.34$): The more favorable ability to recover, overcommitment and emotional exhaustion are, the more probability there is of reaching the regular date of retirement. For age, there is a very small ($r = -0.15$) correlation for male teachers and a small ($r = -0.21$) correlation for female teachers.

Regardless of gender, the trend of increasing weekly working time for teachers is accompanied by a higher overcommitment ($r = 0.26$), reduced ability to recover ($r = 0.26$), and increased ERI ($r = 0.23$), and vice versa; while for emotional exhaustion there is only a very small correlation with weekly working time ($r = 0.17$). At the same time, increasing effort-reward ratio, overcommitment ($r = 0.42$), inability to recover ($r = 0.49$), and emotional exhaustion ($r = 0.44$) are on the rise. Emotional exhaustion is moderately correlated with overcommitment ($r = 0.53$) and inability to recovery ($r = 0.57$). And there is a strong correlation between overcommitment and inability to recover ($r = 0.77$), according to which pronounced overcommitment is associated with strong inability to recover.

Binary logistic regression analyses were calculated separately for both genders to examine the extent to which the characteristics studied contribute to reaching the regular retirement age among male and female teachers. The results of these analyses make clear that the models of the two gender groups practically do not differ. When looking at the individual (independent) characteristics, it turns out that the control variable subject profile and the working time-related characteristics hardly contribute to the explanation of reaching regular retirement age (Nagelkerke $R^2 < 1\%$). The ER model characteristics explain 8 to 13%, and the person-related characteristics 6 to 15% of the probability of reaching regular retirement age. For both gender groups emotional exhaustion (15–17%), ability to recover (11–14%), and the effort-reward ratio (12–13%) provide the highest-related explanations.

Therefore, and based on the correlation analyses, only total weekly working time, ER ratio, the person-related characteristics, and emotional exhaustion were included in the overall model, as well as age as a control variable. The propensity to overcommit and the ability to recover alone explain 11–15% and the addition of emotional exhaustion 17–21% of the variance for attainment of regular retirement age, whereby the correlation between overcommitment and ability to recover should be noted ($r = 0.77$).

The highest variance clarification could be achieved with the overall model (see **Table 5**). For both genders, this model is statistically significant (male/female: $\chi^2(6) = 520.80/705.72$, $p < 0.001$); however, at 25% (Nagelkerke $R^2 = 0.25$), it shows only an acceptable goodness-of-fit between the overall model and the data (Backhaus et al., 2003), which means the independent variables explain 25% of the probability of teachers reaching regular retirement age.

According to the percentage of accuracy in classification, only 36% of the statements of male teachers who cannot imagine reaching regular retirement age (288 out of 791) were predicted correctly. For female teachers, this concerns 54% (774 out of 1,441). In comparison, 92% of the statements

TABLE 5 | Binary logistic regression models of work-, person-, and health-related characteristics and covariates (age and subject profile) with reaching regular retirement of male ($n = 2{,}680$) and female ($n = 3{,}429$) full-time teachers.

Total model	Coefficient (*B*)	Standard error of *B*	Wald statistic	Value of *p*	Estimated odds ratio	Confidence interval for Exp (*B*)	
					Exp (*B*)	Lower limit	Upper limit
Male							
Working time [hours/week]	0.002	0.01	0.17	0.677	1.002	0.99	1.01
Effort-reward ratio	–0.876	0.13	45.62	<0.001***	0.416	0.32	0.54
Overcommitment [pts]	–0.143	0.02	40.80	<0.001***	0.867	0.83	0.91
Inability to recover [pts]	–0.247	0.04	17.61	<0.001***	0.781	0.70	0.88
Emotional exhaustion [pts]	–0.423	0.05	87.82	<0.001***	0.655	0.60	0.72
Age [years]	–0.034	0.00	48.12	<0.001***	0.966	0.96	0.98
Constant	5.441	0.42	170.45	<0.001***	230.784		
Female							
Working time [hours/week]	0.017	0.01	13.99	<0.001***	1.018	1.01	1.03
Effort-reward ratio	–0.860	0.11	580.02	<0.001***	0.423	0.34	0.53
Overcommitment [pts]	–0.17	0.05	11.92	0.010**	0.844	0.77	0.93
Inability to recover [pts]	–0.095	0.02	24.70	<0.001***	0.910	0.88	0.94
Emotional exhaustion [pts]	–0.454	0.04	139.62	<0.001***	0.635	0.59	0.68
Age [years]	–0.041	0.01	118.22	<0.001***	0.960	0.95	0.97
Constant	3.943	0.33	143.51	<0.001***	51.570		

*Dependent variable: early vs. regular retirement = 0–1 coded; binary logistic regressions (method: enter), Exp (B) = expected ß. CI: confidence interval, significance (two-sided): ***p < 0.001, **p < 0.01. Nagelkerke R^2: male = 0.251, female = 0.250. For regression analyses, the problem of collinearity must be taken into account. In the analyses, overcommitment and inability to recover are correlated with r = 0.77. Clear collinearity is accepted for r > 0.90 (e.g., Tabachnick and Fidell, 2013; Harlow, 2014).*

made by male teachers and 81% by female teachers who stated that they intend to retire at the regular retirement start date are correctly assigned. Overall, this corresponds to a correct prediction in 76% of cases for male teachers and in 69% of cases for female teachers. All model coefficients and odds can be found in **Table 5**.

The main predictors for reaching the regular retirement are emotional exhaustion, ER ratio, inability to recover, and overcommitment: As the ER ratio increases and the EE increases, the chance of regular retirement is reduced by a factor of 0.416 and 0.655 for men and by a factor of 0.423 and 0.635 for women, respectively. For both genders, these correlations are weaker for the inability to recover (OR: male = 0.867, female = 0.910) and overcommitment (OR = 0.781 and 0.844, respectively).

DISCUSSION

Teachers regularly perform a wide range of tasks with high psychosocial and emotional demands. In doing so, the high level of autonomy in the work organization represents both a resource and a risk for the long-term health of teachers. If a good balance between work and recovery is achieved, teachers can stay in the profession for a long time. However, if an unfavorable working style leads to high levels of professional exhaustion and an inability to recover, teachers are at increased risk of stress-related mental illness and early retirement. It is therefore of paramount importance to analyze possible factors of influence in this process and to examine whether in different ways males and females endanger their own health due to particular behavior, thereby increasing the probability of early retirement.

Our study shows that the overall working time of full-time high school teachers differs only slightly ($\eta^2_{\text{partial}} < 0.01$). The only difference is that, on average, female teachers spend about 2 h a week more on teaching activities than male teachers ($\eta^2_{\text{partial}} = 0.016$, small effect). It is noticeable that the extent of emotional exhaustion does not differ between the genders ($\eta^2_{\text{partial}} < 0.01$). Previous studies came to contradictory results (e.g., Bakker et al., 2002; Bekker et al., 2005; Rupert and Morgan, 2005). In some studies, female teachers have reported higher levels of emotional exhaustion than male teachers (Van Dick and Wagner, 2001; Wang et al., 2015; Arvidsson et al., 2016). A frequent justification for this has been the assumed greater overall burden on women due to increased family obligations and work-family conflicts (Gyllensten and Palmer, 2005). Since this study only looked at full-time teachers and noticeably few female teachers cared for children in their own household (21%), the lack of gender differences seems plausible. In addition, family responsibilities are now shared more fairly between men and women than they were 20 years ago.

It should be noted that in total around one-quarter of male teachers (23%) and one-third of female teachers (31%) have high levels of emotional exhaustion. This is a disturbing finding since it is evident that high exhaustion values are an important risk factor in deciding on early retirement or for moving to other professions (Alarcon, 2011; Van Droogenbroeck and Spruyt, 2014). Moreover, emotional exhaustion is closely linked

to both work satisfaction (Klusmann et al., 2008; Skaalvik and Skaalvik, 2010) and student achievement (Klusmann et al., 2016).

In terms of personal traits, the results show significant differences in self-harming behavior between male and female teachers. For example, significantly more female teachers (35%) than male teachers (21%) are affected by a high level of overcommitment and a high inability to recover. These teachers are at increased risk to their health in the medium and long term, especially from exhaustion. Only half of teachers (51%) have normal values for both overcommitment and ability to recover (men: 59% and women: 44%).

The regression models confirm that these personal characteristics contribute to the prediction of early retirement. Emotional exhaustion, inability to recover, overcommitment, and ER ratio are identified as important predictors for entering retirement for both genders. The results imply an increase in the probability of early retirement with increasing emotional exhaustion and inability to recover, and a high level of overcommitment. Similarly, an increase in the imbalance between effort and reward (ER ratio) increases the probability of leaving the profession early. Age effects tend to be subordinate in both genders, with the slight trend that as teachers age, they are more likely to anticipate retirement. Considerably more female teachers (42%) than male teachers (30%) predict early retirement ($d = 261$, small effect).

However, the two regression models have a low sensitivity and account for only 25% of the variation in each gender group. This means that retirement age is affected by other features not studied here. Van Droogenbroeck and Spruyt (2014), in a sample of more than 3,000 Belgian teachers, 60% of whom were in employment and 40% were already retired, identified gender, emotional exhaustion, and financial security (e.g., own property) among others as significant predictors of a retirement decision. As in this study, female Belgian teachers want to retire more than their male colleagues. Irrespective of gender, two-thirds of teachers want and make use of early retirement. Teacher turnover is a long-known phenomenon in the teaching profession, with the highest dropouts in early and late career (Grissmer and Kirby, 1997). Harris and Adams (2007) showed that the early retirement of teachers is a particular issue in comparison with other professions (including nurses and social workers).

Although there are many reasons for leaving the profession in this study, it is surprising that there is a strong match between male and female teachers: high workload is the main reason for early retirement for more than three quarters (79%) of all teachers. However, excessive work demands are not only perceived by teachers subjectively (Bauer et al., 2007; Skaalvik and Skaalvik, 2010), but also are considered the most important cause of stress (Kyriacou, 2001) and reduced wellbeing (Skaalvik and Skaalvik, 2017b) in the teaching profession. In addition, high work demands have been proven to be related to emotional exhaustion (Antoniou et al., 2006; Hakanen et al., 2006; Skaalvik and Skaalvik, 2017a; Baeriswyl et al., 2021).

Whether high work requirements ultimately become a health risk depends on the working conditions themselves and how an individual deals with these requirements. The effort-reward model (Siegrist et al., 2009) contains both explanatory approaches to the relationship between work requirements and health, including the intrinsic feature overcommitment. Both ERI (ER ratio > 1) as well as an excess tendency to overcommit were identified as predictors for early retirement. For the ER ratio, there was no gender effect, but for the overcommitment there was. It is worth noting that over a third (35%) of all teachers surveyed reported an ERI. In contrast, the share of ERI was significantly lower (22%) in a previous study among German teachers by Unterbrink et al. (2007). They also showed no gender effects in the ER ratio; however, they found an age effect. Teachers aged 45 and over reported higher ER ratios than their younger colleagues. Hinz et al. (2016), in contrast, in a recent German study, stated a slightly lower ER ratio for female teachers (0.63) than for male teachers (0.69; $\eta^2_{\text{partial}} = 0.11$, medium effect), without evidence of a significant age effect. Although the current gender impact study appears inconsistent. Siegrist (2017) drew attention to the high prevalence of ERI in education and showed a link between ERI, exhaustion, and depression in teachers.

Overcommitment may also increases the risk of exhaustion (Bakker et al., 2000; Wang et al., 2015). Our results show significant gender differences for overcommitment. Almost half of female teachers (47%) and at least one-third (33%) of male teachers are excessively likely to overcommit. While the direct health effect of overcommitment has been demonstrated robustly, it has not yet been fully clarified whether overcommitment further moderates the relationship between effort and reward (Siegrist and Li, 2016).

The second model looked at coping patterns, the inability of recovery, and it also showed that male and female teachers cope differently with their professional needs. For example, insufficient recovery in our sample was significantly more common among female (47%) than among male teachers (38%; $\eta^2_{\text{partial}} = 0.033$, small effect). The argument that family responsibilities could lead to an inability for recovery is not true in our sample as a justification for gender differences, because only one in five female teachers looked after children in their own household. But it may be that female teachers find it harder to mentally detach from work content. Some studies have shown that people who tend to overcommit have a lower ability to switch off from work (Feldt et al., 2013; Wendsche and Lohmann-Haislah, 2017).

Mental detachment is an essential condition for recovery (Sonnentag and Fritz, 2015). This seems to be a particular problem in the teaching profession. Varol et al. (2021) showed, on the basis of a representative survey of German employees, that teachers report difficulties in switching off from work mentally twice as often compared to other professions (42% vs. 21%). Emotional requirements as well as time and performance pressure were the main causes of this. Overall, teachers were the second most frequently (23%) affected by recovery problems after managers (Schulz et al., 2020).

For teachers, recovery after work is particularly important, as their opportunities for rest at work are insufficient (Geurts, 2014). In addition, unfavorable working hours in the evening and the weekends hinder necessary recovery processes due to consistent physiological activation (Van der Hulst, 2003; Sonnentag and Fritz, 2015) and may stimulate rumination.

Ruminating, which females have a stronger propensity for than males (Jose and Brown, 2008; Hyde, 2014), can continue the process of not being able to switch off. In a recent meta-analysis, Karabinski et al. (2021) showed that a variety of interventions can effectively support detaching from work, especially when the programs use either boundary management strategies, emotional regulation techniques, or strategies to improve sleep quality as key elements. Interventions with higher intensity and longer duration achieved the greatest success. Older participants and those with health impairments benefited more from the programs.

In summary, the present study shows that for full-time high school teachers, the work-related characteristics are not differ between the genders ($\eta^2_{\text{partial}} < 0.01$). Even for emotional exhaustion, the differences between the two gender groups are not relevant. On the other hand, the personal characteristics of overcommitment and inability to recover are significantly less favorable for female teachers than for male teachers. As both behavioral characteristics are considered ineffective coping strategies and they further increase the stress caused by working conditions, there is evidence of more self-harming behavior among female teachers than among their male counterparts. The assumption is supported by the significantly more frequent perception of females that they cannot remain in the profession until the regular retirement age.

The originality of the study is that for the first time, data on working time, work load, and health are reported with a large and representative sample of full-time high school teachers for the whole of Germany, taking into account significant influence factors. The composition of the sample corresponds to the characteristics of gender and age of the German high school teaching population. These represent a large professional group, which throughout Germany comprises approximately 42% men and 58% women, thus allowing good comparability for the consequences of gender-specific, occupational stress (Travers and Cooper, 1991).

Another feature of this study is the relatively homogeneous sample, which only takes into account full-time upper-level high school teachers for whom the share of teaching dominates; teachers in management positions (e.g., school directors) and officials (e.g., staff councils) were consistently excluded. Both genders have comparable working conditions, which is considered a crucial prerequisite to detect gender effects (Schaufeli et al., 2001). Previous research has often looked at inhomogeneous samples and reported gender differences with no indication of effect sizes (Bauer et al., 2006; Unterbrink et al., 2007; Nübling et al., 2011; Hinz et al., 2016).

From a methodological point of view, the study is also based on differentiated working time records with 12 categories of activities over 4 weeks (online protocol). Even if this period represents only an average workload from the school year, this method of collection provides a reliable basis for determining average weekly working time (Felsing et al., 2019b). In addition, answers relating to early retirement and possible measures to achieve normal retirement age were evaluated with great effort for all 6,109 teachers and categorized according to self-developed categories. As teachers are considered experts in this context, a differentiated picture of their occupational and health sources and resources could be generated.

There are also limitations to consider when interpreting the results in this study. The data were collected as a cross-sectional study, so that it is not possible to tell causal links between the characteristics examined and the projected retirement age of the teachers. Since participation in the investigation was voluntary, it is also a convenience sample in which selection effects and a healthy worker effect cannot be excluded. As a result, health risks may have been underestimated.

Another limitation concerns data collection: since the variables were captured by self-information, known bias due to social desirability, response tendencies and memory deficits cannot be excluded. Furthermore, the probability of retirement was collected only as a single item. As the focus on this issue is on content validity, this method of survey is appropriate (Fisher et al., 2016b). According to de Boer et al. (2004), a single global issue of validity and reliability does not have to have significant disadvantages over larger sets of questions.

With regard to the regression analyses used, it should be noted that ultimately both gender and the confounders adjusted in the statistical analyses could be partially related to early retirement. In order to achieve more clarity about the influence of the variables, alternative analyses, such as a propensity score matching analyses, could be applied in the future.

CONCLUSION

The study supports the known findings that teachers need more support to stay healthy. This is the most important prerequisite for dealing with the demanding work requirements, to remain efficient and to provide good teaching. The sample in this study provides a solid basis to derive proportionate and behavioral prevention measures. The key is to identify health risks at an early stage and to influence labor and health resources in such a way as to counteract widespread premature retirement among teachers.

In order to protect against overcommitment and health problems, measures are needed which focus on reducing work requirements and developing social support work environments, while at the same time focusing on individual improvement of coping strategies in dealing with the high workloads and emotional interactions. The teachers in our study themselves propose reducing the class size and the number of compulsory hours, reducing additional tasks, bureaucratic structures and administrative burdens, and above all improving the organizational and working conditions in schools. This includes providing high-quality teaching materials and creating adequate retreats at school, as well as a value-added management style and team-oriented approach among all employees. Teachers also want more realistic curricula and more time to maintain relationships with students. In addition, in the teaching profession, there is a lack of well-founded and proven human resources development strategies which maintain and promote the employability and health of teachers until their regular retirement. It would be advisable to offer preventive medical care on a regular basis, covering early indicators of

health risk, such as overcommitment, inability to recover, and emotional exhaustion, and use them as a basis for individual health advice for teachers.

On an individual level, aspiring teachers should learn in their studies techniques and strategies which contribute to the regeneration and strengthening of resilience and which can be integrated into both professional and private everyday life. These includes active recreation offers, such as activities in nature, which help to switch off from work. It is also necessary for teachers to develop a healthy distance from the many requirements of the teaching profession.

In the future, it will be essential to have longitudinal studies to analyze the links between school workload and health consequences, as well as health prevention.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article materials, and further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the University of Rostock (A 2018-0031). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RS made the funding acquisition. SK and RS designed the study, made the project administration, collected the data, made the analysis and interpretation of the data, and wrote the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

This study was funded from the German Association of Philologists (DPhV). The sponsor was not involved in data collection, analysis, interpretation, manuscript preparation, or the decision to publish.

ACKNOWLEDGMENTS

The scientists are dedicated to Christine Petzke who helped categorize the reasons for early retirement and the proposed measures and to all the teachers participated in the research project.

REFERENCES

Agaiâ-Demjaha, T., Minov, J., Stoleski, S., and Zafirova, B. (2015). Stress causing factors among teachers in elementary schools and their relationship with demographic and job characteristics. *Open Access Maced. J. Med. Sci.* 3, 493–499. doi: 10.3889/oamjms.2015.077

Alarcon, G. M. (2011). A meta-analysis of burnout with job demands, resources, and attitudes. *J. Vocat. Behav.* 79, 549–562. doi: 10.1016/j.jvb.2011.03.007

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., and Trautwein, U. (2018). Student misbehavior and teacher well-being: testing the mediating role of the teacher-student relationship. *Learn. Instr.* 58, 126–136. doi: 10.1016/j.learninstruc.2018.05.006

Antoniou, A. S., Polychroni, F., and Vlachakis, A. N. (2006). Gender and age differences in occupational stress and professional burnout between primary and high-school teachers in Greece. *J. Manag. Psychol.* 21, 682–690. doi: 10.1108/02683940610690213

Arntén, A.-C. A., Jansson, B., and Archer, T. (2008). Influence of affective personality type and gender upon coping behavior, mood and stress. *Individ. Differ. Res.* 6, 139–168.

Arroba, T., and James, K. (2002). *Pressure at Work: A Survival Guide*. Maidenhead, UK: McGraw-Hill.

Arvidsson, I., Håkansson, C., Karlson, B., Björk, J., and Persson, R. (2016). Burnout among Swedish school teachers – a cross-sectional analysis. *BMC Public Health* 16:823. doi: 10.1186/s12889-016-3498-7

Ashforth, B. E., Kreiner, G. E., and Fugate, M. (2000). All in a day's work: boundaries and micro role transitions. *Acad. Manag. Rev.* 25, 472–491. doi: 10.2307/259305

Backhaus, K., Erichson, B., Plinke, W., and Weiber, R. (2003). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Berlin: Springer.

Baeriswyl, S., Bratoljic, C., and Krause, A. (2021). How homeroom teachers cope with high demands: effect of prolonging working hours on emotional exhaustion. *J. Sch. Psychol.* 85, 125–139. doi: 10.1016/j.jsp.2021.02.002

Bakker, A. B., Demerouti, E., and Schaufeli, W. B. (2002). Validation of the Maslach Burnout Inventory - general survey: an internet study. *Anxiety Stress Coping* 15, 245–260. doi: 10.1080/1061580021000020716

Bakker, A. B., Killmer, C. H., Siegrist, J., and Schaufeli, W. B. (2000). Effort–reward imbalance and burnout among nurses. *J. Adv. Nurs.* 31, 884–891. doi: 10.1046/j.1365-2648.2000.01361.x

Bauer, J., Stamm, A., Virnich, K., Wissing, K., Müller, U., Wirsching, M., et al. (2006). Correlation between burnout syndrome and psychological and psychosomatic symptoms among teachers. *Int. Arch. Occup. Environ. Health* 79, 199–204. doi: 10.1007/s00420-005-0050-y

Bauer, J., Unterbrink, T., Hack, A., Pfeifer, R., Buhl-Grießhaber, V., Müller, U., et al. (2007). Working conditions, adverse events and mental health problems in a sample of 949 German teachers. *Int. Arch. Occup. Environ. Health* 80, 442–449. doi: 10.1007/s00420-007-0170-7

Beehr, T. A. (1986). The process of retirement: a review and recommendations for future investigation. *Pers. Psychol.* 39, 31–55. doi: 10.1111/j.1744-6570.1986.tb00573.x

Bekker, M. H. J., Croon, M. A., and Bressers, B. (2005). Childcare involvement, job characteristics, gender and work attitudes as predictors of emotional exhaustion and sickness absence. *Work Stress* 19, 221–237. doi: 10.1080/02678370500286095

Blanz, M. (ed.) (2015). "Gütekriterien von Testverfahren," in *Forschungsmethoden und Statistik für die Soziale Arbeit: Grundlagen und Anwendungen*. Stuttgart: Kohlhammer, 255–259.

Bond, M. A., Punnett, L., Pyle, J. L., Cazeca, D., and Cooperman, M. (2004). Gendered work conditions, health, and work outcomes. *J. Occup. Health Psychol.* 9, 28–45. doi: 10.1037/1076-8998.9.1.28

Borg, M. G. (1990). Occupational stress in British educational settings: a review. *Educ. Psychol.* 10, 103–126. doi: 10.1080/0144341900100201

Bühl, A. (2016). *SPSS 23. Einführung in die moderne Datenanalyse*. Hallbergmoos: Pearson.

Chaplain, R. P. (2008). Stress and psychological distress among trainee secondary teachers in England. *Educ. Psychol.* 28, 195–209. doi: 10.1080/01443410701491858

Chong, E. Y. L., and Chan, A. H. S. (2010). Subjective health complaints of teachers from primary and secondary schools in Hong Kong. *Int. J. Occup. Saf. Ergon.* 16, 23–39. doi: 10.1080/10803548.2010.11076825

Clark, S. C. (2000). Work/family border theory: a new theory of work/family balance. *Hum. Relat.* 53, 747–770. doi: 10.1177/0018726700536001

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cropley, M., and Zijlstra, F. R. H. (2011). "Work and rumination," in *Handbook of Stress in the Occupations*. eds. J. Langan-Fox and C. L. Cooper (UK: Edward Elgar Publishing Ltd), 487–503.

De Boer, A. G. E. M., van Lanschot, J. J. B., Stalmeier, P. F. M., van Sandick, J. W., Hulscher, J. B. F., de Haes, J., et al. (2004). Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Qual. Life Res.* 13, 311–320. doi: 10.1023/B:QURE.0000018499.64574.1f

Dettmers, J., Deci, N., Baeriswyl, S., Berset, M., and Krause, A. (2016). "Self-endangering work behavior," in *Healthy at Work: Interdisciplinary Perspectives*. eds. M. Wiencke, M. Cacace and S. Fischer (Cham: Springer International Publishing), 37–51.

European Commission/Eurydice (2013). *Key Data on Teachers and School Leaders in Europe*. Luxembourg: Publications Office of the European Union.

Feldt, T., Huhtala, M., Kinnunen, U., Hyvönen, K., Mäkikangas, A., and Sonnentag, S. (2013). Long-term patterns of effort-reward imbalance and over-commitment: investigating occupational well-being and recovery experiences as outcomes. *Work Stress.* 27, 64–87. doi: 10.1080/02678373.2013.765670

Felsing, C., Kreuzfeld, S., Stoll, R., and Seibt, R. (2019a). App-basierte vs. geschätzte Ermittlung der Arbeitszeit von Gymnasiallehrkräften. *Präv Gesundheitsförd* 14, 281–289. doi: 10.1007/s11553-018-0682-x

Felsing, C., Seibt, R., Stoll, R., and Kreuzfeld, S. (2019b). Working time structure of secondary school teachers in the course of the day and the week – an app-based pilot study. *Arbeitsmed. Sozialmed. Umweltmed.* 54, 47–55.

Fisher, G. G., Chaffee, D. S., and Sonnega, A. (2016a). Retirement timing: a review and recommendations for future research. *Work Aging Retire.* 2, 230–261. doi: 10.1093/workar/waw001

Fisher, G. G., Matthews, R. A., and Gibbons, A. M. (2016b). Developing and investigating the use of single-item measures in organizational research. *J. Occup. Health Psychol.* 21, 3–23. doi: 10.1037/a0039139

Folkman, S., Lazarus, R. S., Dunkel-Schetter, C., Delongis, A., and Gruen, R. J. (1986). Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes. *J. Pers. Soc. Psychol.* 50, 992–1003. doi: 10.1037/0022-3514.50.5.992

Framke, E., Sørensen, J. K., Alexanderson, K., Farrants, K., Kivimäki, M., Nyberg, S. T., et al. (2021). Emotional demands at work and risk of long-term sickness absence in 1·5 million employees in Denmark: a prospective cohort study on effect modifiers. *Lancet Public Health* 6, 752–759. doi: 10.1016/S2468-2667(21)00185-7

Frankenhaeuser, M. (1991). The psychophysiology of workload, stress, and health: comparison between the sexes. *Ann. Behav. Med.* 13, 197–204.

García-Carmona, M., Marín, M. D., and Aguayo, R. (2019). Burnout syndrome in secondary school teachers: a systematic review and meta-analysis. *Soc. Psychol. Educ.* 22, 189–208. doi: 10.1007/s11218-018-9471-9

Geurts, S. A. E. (2014). "Recovery from work during off-job time," in *Bridging Occupational, Organizational and Public Health: A Transdisciplinary Approach*. eds. G. F. Bauer and O. Hämmig (Cham: Springer Science + Business Media), 193–208.

Greenglass, E. R., and Burke, R. J. (2003). "Teacher stress," in *Occupational Stress in the Service Professions*. eds. M. F. Dollard and A. H. Winefield (New York, NY: Taylor and Francis), 213–236.

Griffith, J., Steptoe, A., and Cropley, M. (1999). An investigation of coping strategies associated with job stress in teachers. *Br. J. Educ. Psychol.* 69, 517–531. doi: 10.1348/000709999157879

Grissmer, W. D., and Kirby, S. N. (1997). Teacher turnover and teacher quality. *Teach. Coll. Rec.* 99, 45–56.

Guglielmi, R. S., and Tatrow, K. (1998). Occupational stress, burnout, and health in teachers: a methodological and theoretical analysis. *Rev. Educ. Res.* 68, 61–99. doi: 10.3102/00346543068001061

Gyllensten, K., and Palmer, S. (2005). The role of gender in workplace stress: a critical literature review. *Health Educ. J.* 64, 271–288. doi: 10.1177/0017896905064003

Hakanen, J. J., Bakker, A. B., and Schaufeli, W. B. (2006). Burnout and work engagement among teachers. *J. Sch. Psychol.* 43, 495–513. doi: 10.1016/j.jsp.2005.11.001

Harlow, L. L. (2014). *The Essence of Multivariate Thinking: Basic Themes and Methods. 2nd Edn*. New York, NY: Routledge/Taylor & Francis Group.

Harris, D. N., and Adams, S. J. (2007). Understanding the level and causes of teacher turnover: a comparison with other professions. *Econ. Educ. Rev.* 26, 325–337. doi: 10.1016/j.econedurev.2005.09.007

Hinsch, D. M., Spanier, K., Radoschewski, F. M., and Bethge, M. (2019). Associations between overcommitment, effort–reward imbalance and mental health: findings from a longitudinal study. *Int. Arch. Occup. Environ. Health* 92, 559–567. doi: 10.1007/s00420-018-1391-7

Hinz, A., Zenger, M., Brähler, E., Spitzer, S., Scheuch, K., and Seibt, R. (2016). Effort–reward imbalance and mental health problems in 1074 German teachers, compared with those in the general population. *Stress. Health* 32, 224–230. doi: 10.1002/smi.2596

Höge, T., and Hornung, S. (2013). Perceived flexibility requirements: exploring mediating mechanisms in positive and negative effects on worker well-being. *Econ. Ind. Democracy* 36, 407–430. doi: 10.1177/0143831X13511274

Hyde, J. S. (2014). Gender similarities and differences. *Annu. Rev. Psychol.* 65, 373–398. doi: 10.1146/annurev-psych-010213-115057

Jepson, E., and Forrest, S. (2006). Individual contributory factors in teacher stress: the role of achievement striving and occupational commitment. *Br. J. Educ. Psychol.* 76, 183–197. doi: 10.1348/000709905X37299

Jose, P. E., and Brown, I. (2008). When does the gender difference in rumination begin? Gender and age differences in the use of rumination by adolescents. *J. Youth Adolesc.* 37, 180–192. doi: 10.1007/s10964-006-9166-y

Karabinski, T., Haun, V. C., Nübold, A., Wendsche, J., and Wegge, J. (2021). Interventions for improving psychological detachment from work: a meta-analysis. *J. Occup. Health Psychol.* 26, 224–242. doi: 10.1037/ocp0000280

Klassen, R. M., Perry, N. E., and Frenzel, A. C. (2012). Teachers' relatedness with students: an underemphasized component of teachers' basic psychological needs. *J. Educ. Psychol.* 104, 150–165. doi: 10.1037/a0026253

Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., and Baumert, J. (2008). Teachers' occupational well-being and quality of instruction: the important role of self-regulatory patterns. *J. Educ. Psychol.* 100, 702–715. doi: 10.1037/0022-0663.100.3.702

Klusmann, U., Richter, D., and Lüdtke, O. (2016). Teachers' emotional exhaustion is negatively related to students' achievement: evidence from a large-scale assessment study. *J. Educ. Psychol.* 108, 1193–1203. doi: 10.1037/edu0000125

Krause, A., Berset, M., and Peters, K. (2015). Interested self-endangerment – from the direct to indirect control [Interssierte Selbstgefährdung – von der direkten zur indirekten Steuerung]. *Arbeitsmed. Sozialmed. Umweltmed.* 50, 164–170.

Kyriacou, C. (2001). Teacher stress: directions for future research. *Educ. Rev.* 53, 27–35. doi: 10.1080/00131910120033628

Lagrosen, Y., and Lagrosen, S. (2020). Gender, quality and health – a study of Swedish secondary school teachers. *Int. J. Workplace Health Manag.* 13, 223–238. doi: 10.1108/IJWHM-09-2018-0125

Laughlin, A. (1984). Teacher stress in an Australian setting: the role of biographical mediators. *Educ. Stud.* 10, 7–22. doi: 10.1080/0305569840100102

Lehr, D., Hillert, A., and Keller, S. (2009). What can balance the effort? Associations between effort-reward imbalance, overcommitment, and affective disorders in German teachers. *Int. J. Occup. Environ. Health* 15, 374–384. doi: 10.1348/096317909X406772

Leung, D. Y. P., and Lee, W. W. S. (2006). Predicting intention to quit among Chinese teachers: differential predictability of the components of burnout. *Anxiety Stress Coping* 19, 129–141. doi: 10.1080/10615800600565476

Li, J., Yang, W., and Cho, S.-I. (2006). Gender differences in job strain, effort-reward imbalance, and health functioning among Chinese physicians. *Soc. Sci. Med.* 62, 1066–1077. doi: 10.1016/j.socscimed.2005.07.011

Liu, C., Spector, P. E., and Shi, L. (2008). Use of both qualitative and quantitative approaches to study job stress in different gender and occupational groups. *J. Occup. Health Psychol.* 413, 357–370. doi: 10.1037/1076-8998.13.4.357

Loerbroks, A., Meng, H., Chen, M.-L., Herr, R., Angerer, P., and Li, J. (2014). Primary school teachers in China: associations of organizational justice and effort–reward imbalance with burnout and intentions to leave the profession in a cross-sectional sample. *Int. Arch. Occup. Environ. Health* 87, 695–703. doi: 10.1007/s00420-013-0912-7

Madsen, I. E. H., and Rugulies, R. (2021). Understanding the impact of psychosocial working conditions on workers' health: we have come a long way, but are we there yet? *Scand. J. Work Environ. Health* 47, 483–487. doi: 10.5271/sjweh.3984

Maslach, C., and Jackson, S. E. (1981). The measurement of experienced burnout. *J. Organ. Behav.* 2, 99–113. doi: 10.1002/job.4030020205

Maslach, C., and Jackson, S.E. (1986). *Maslach Burnout Inventory: Manual*. Palo Alto, CA: Consulting Psychologists Press.

Meijman, T.F., and Mulder, G. (1998). "Psychological aspects of workload," in *Handbook of Work and Organizational Psychology*. eds. P. J. D. Drenth, H. Thierry and C. J. De Wolff (Hove, UK: Psychology Press/ Erlbaum), 5–33.

Niedhammer, I., Sultan-Taïeb, H., Chastang, J.-F., Vermeylen, G., and Parent-Thirion, A. (2014). Fractions of cardiovascular diseases and mental disorders attributable to psychosocial work factors in 31 countries in Europe. *Int. Arch. Occup. Environ. Health* 87, 403–411. doi: 10.1007/s00420-013-0879-4

Nübling, M., Vomstein, M., Haug, A. T. N., and Adiwidjaja, A. (2011). "European-wide survey on teachers work related stress –assessment, comparison and evaluation of the impact of psychosocial hazards on teachers at their workplace". https://www.etuce.homestead.com/Publications2011/Final_Report_on_the_survey_on_WRS-2011-eng.pdf (Accessed December 1, 2021).

OECD (2019). *Education at a Glance 2019: OECD Indicators*. Paris: OECD Publishing.

Persson, R., Hansen, Å.-M., Ohlsson, K., Balogh, I., Nordander, C., and Ørbæk, P. (2008). Physiological and psychological reactions to work in men and women with identical job tasks. *Eur. J. App. Physiol.* 105, 595–606. doi: 10.1007/s00421-008-0939-8

Rasku, A., and Kinnunen, U. (2003). Job conditions and wellness among Finnish upper secondary school teachers. *Psychol. Health* 18, 441–456. doi: 10.1080/0887044031000147184

Rau, R., and Triemer, A. (2004). Overtime in relation to blood pressure and mood during work, leisure, and night time. *Soc. Indic. Res.* 67, 51–73. doi: 10.1023/B:SOCI.0000007334.20490.52

Reilly, E., Dhingra, K., and Boduszek, D. (2014). Teachers' self-efficacy beliefs, self-esteem, and job stress as determinants of job satisfaction. *Int. J. Educ. Manag.* 28, 365–378. doi: 10.1108/IJEM-04-2013-0053

Richter, P., Hille, B., and Rudolf, M. (1999). Gesundheitsrelevante Bewältigung von Arbeitsanforderungen. *Z Differ Diagn. Psychol.* 20, 25–38. doi: 10.1024//0170-1789.20.1.25

Richter, P., Rotheiler, E., and Rudolf, M. (2015). *FABA – Fragebogen zur Analyse belastungsrelevanter Anforderungsbewältigung*. Mattersburg (Austria): P & T Prieler Tometich Verlag: Potentialanalyse und Testverfahren GmbH.

Richter, P., Rudolf, M., and Schmidt, C. F. (1996). *FABA: Fragebogen zur Erfassung beanspruchungsrelevanter Anforderungsbewältigung*. Frankfurt: Harcourt Test Service.

Roxburgh, S. (1996). Gender differences in work and well-being: effects of exposure and vulnerability. *J. Health Soc. Behav.* 37, 265–277. doi: 10.2307/2137296

Rubino, C., Volpone, S. D., and Avery, D. R. (2013). Burnout on mars and venus: exploring gender differences in emotional exhaustion. *Gender Manage.* 28, 74–93. doi: 10.1108/17542411311303220

Rugulies, R., Aust, B., and Madsen, I. E. H. (2017). Effort–reward imbalance at work and risk of depressive disorders. A systematic review and meta-analysis of prospective cohort studies. *Scand. J. Work Environ. Health* 43, 294–306. doi: 10.5271/sjweh.3632

Rupert, P. A., and Morgan, D. J. (2005). Work setting and burnout among professional psychologists. *Prof. Psychol. Res. Pract.* 36, 544–550. doi: 10.1037/0735-7028.36.5.544

Schaufeli, W. B., Bakker, A. B., Hoogduin, K., Schaap, C., and Kladler, A. (2001). On the clinical validity of the Maslach Burnout Inventory and the burnout measure. *Psychol. Health* 16, 565–582. doi: 10.1080/08870440108405527

Schaufeli, W. B., Leiter, M. P., Maslach, C., and Jackson, S. E. (1996). "Maslach Burnout Inventory – general survey," in *Maslach Burnout Inventory Manual*. eds. C. Maslach, S. E. Jackson and M. P. Leiter (Palo Alto, CA: Consulting Psychologists Press), 19–26.

Schult, J., Münzer-Schrobildgen, M., and Sparfeldt, J. R. (2014). Belastet, aber hochzufrieden? Arbeitsbelastung von Lehrkräften im Quer- und Längsschnitt. *Z Gesundheitspsychol* 22, 61–67. doi: 10.1026/0943-8149/a000114

Schulz, A. D., Wendsche, J., Lohmann-Haislah, A., and Schöllgen, I. (2020). Erholungsbeeinträchtigungen bei Beschäftigten. *Zbl. Arbeitsmed.* 70, 57–65. doi: 10.1007/s40664-019-00373-7

Seibt, R., and Kreuzfeld, S. (2021). Influence of work-related and personal characteristics on the burnout risk among full- and part-time teachers. *Int. J. Environ. Res. Public Health* 18:1535. doi: 10.3390/ijerph18041535

Shirom, A., Oliver, A., and Stein, E. H. D. O. A. (2009). Teachers' stressors and strains: a longitudinal study of their relationships. *Int. J. Stress. Manag.* 16, 312–332. doi: 10.1037/a0016842

Siegrist, J. (2017). "Applying occupational health theories to educational stress and health: evidence from the effort-reward imbalance model," in *Educator Stress: An Occupational Health Perspective*. eds. T. M. Mcintyre, S. E. Mcintyre and D. J. Francis (Cham: Springer International Publishing AG), 223–235.

Siegrist, J., and Li, J. (2016). Associations of extrinsic and intrinsic components of work stress with health: a systematic review of evidence on the effort-reward imbalance model. *Int. J. Environ. Res. Public Health* 13:432. doi: 10.3390/ijerph13040432

Siegrist, J., Starke, D., Chandola, T., Godin, I., Marmot, M., Niedhammer, I., et al. (2004). The measurement of effort–reward imbalance at work: European comparisons. *Soc. Sci. Med.* 58, 1483–1499. doi: 10.1016/S0277-9536(03)00351-4

Siegrist, J., Wege, N., Pühlhofer, F., and Wahrendorf, M. (2009). A short generic measure of work stress in the era of globalization: effort–reward imbalance. *Int. Arch. Occup. Environ. Health* 82, 1005–1013. doi: 10.1007/s00420-008-0384-3

Skaalvik, E. M., and Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: a study of relations. *Teach. Teach. Educ.* 26, 1059–1069. doi: 10.1016/j.tate.2009.11.001

Skaalvik, E. M., and Skaalvik, S. (2011a). Teacher job satisfaction and motivation to leave the teaching profession: relations with school context, feeling of belonging, and emotional exhaustion. *Teach. Teach. Educ.* 27, 1029–1038. doi: 10.1016/j.tate.2011.04.001

Skaalvik, E. M., and Skaalvik, S. (2011b). Teachers' feeling of belonging, exhaustion, and job satisfaction: the role of school goal structure and value consonance. *Anxiety Stress Coping* 24, 369–385. doi: 10.1080/z10615806.2010.544300

Skaalvik, E. M., and Skaalvik, S. (2017a). Dimensions of teacher burnout: relations with potential stressors at school. *Soc. Psychol. Educ.* 20, 775–790. doi: 10.1007/s11218-017-9391-0

Skaalvik, E. M., and Skaalvik, S. (2017b). Still motivated to teach? A study of school context variables, stress and job satisfaction among teachers in senior high school. *Soc. Psychol. Educ.* 20, 15–37. doi: 10.1007/s11218-016-9363-9

Solis-Soto, M. T., Schön, A., Parra, M., and Radon, K. (2019). Associations between effort–reward imbalance and health indicators among school teachers in Chuquisaca, Bolivia: a cross-sectional study. *BMJ Open* 9:e025121. doi: 10.1136/bmjopen-2018-025121

Sonnentag, S. (2003). Recovery, work engagement, and proactive behavior: a new look at the interface between nonwork and work. *J. Appl. Psychol.* 88, 518–528. doi: 10.1037/0021-9010.88.3.518

Sonnentag, S., Binnewies, C., and Mojza, E. J. (2010). Staying well and engaged when demands are high: the role of psychological detachment. *J. Appl. Psychol.* 95, 965–976. doi: 10.1037/a0020032

Sonnentag, S., and Fritz, C. (2015). Recovery from job stress: the stressor-detachment model as an integrative framework. *J. Organ. Behav.* 36, 72–103. doi: 10.1002/job.1924

Statistisches Bundesamt (2018). Zahl der Pensionierungen von Lehrkräften 2017 erneut rückläufig. Pressemitteilung Nr. 509 vom 20. Dezember 2018. Available at: https://www.destatis.de/DE/Presse/Pressemitteilungen/2018/12/PD18_509_42.html;jsessionid=A60AFCBFAA01AE32EFB461938EA17717.InternetLive1 (Accessed December 4, 2021).

Stieler, L., Hunger, B., Rudolf, M., Kreuzfeld, S., Stoll, R., and Seibt, R. (2019). Cardiovascular recovery after workload in German shift workers in the hotel and catering industry: results of 24-hour-ambulatory monitoring of heart rate and blood pressure. *Int. Arch. Occup. Environ. Health* 92, 81–91. doi: 10.1007/s00420-018-1357-9

Sünbül, A. M. (2003). An analysis of relations among locus of control, burnout and job satisfaction in Turkish high school teachers. *Aust. J. Educ.* 47, 58–72. doi: 10.1177/000494410304700105

Tabachnick, B. G., and Fidell, L. S. (2013). *Using Multivariate Statistics. 6th Edn*. London: Pearson.

Tamres, L. K., Janicki, D., and Helgeson, V. S. (2002). Sex differences in coping behavior: a meta-analytic review and an examination of relative coping. *Personal. Soc. Psychol. Rev.* 6, 2–30. doi: 10.1207/S15327957PSPR0601_1

Topa, G., Depolo, M., and Alcover, C.-M. (2018). Early retirement: a meta-analysis of its antecedent and subsequent correlates. *Front. Psychol.* 8:2157. doi: 10.3389/fpsyg.2017.02157

Topa, G., Moriano, J. A., Depolo, M., Alcover, C.-M., and Morales, J. F. (2009). Antecedents and consequences of retirement planning and decision-making: a meta-analysis and model. *J. Vocat. Behav.* 75, 38–55. doi: 10.1016/j.jvb.2009.03.002

Travers, C. J., and Cooper, C. L. (1991). Stress and status in teaching: an investigation of potential gender-related relationships. *Women Manag. Rev.* 6, 16–23. doi: 10.1108/09649429110000555

Unterbrink, T., Hack, A., Pfeifer, R., Buhl-Grießhaber, V., Müller, U., Wesche, H., et al. (2007). Burnout and effort–reward-imbalance in a sample of 949 German teachers. *Int. Arch. Occup. Environ. Health* 80, 433–441. doi: 10.1007/s00420-007-0169-0

Van der Hulst, M. (2003). Long workhours and health. *Scand. J. Work Environ. Health* 29, 171–188. doi: 10.5271/sjweh.720

Van Dick, R., and Wagner, U. (2001). Stress and strain in teaching: a structural equation approach. *Br. J. Educ. Psychol.* 71, 243–259. doi: 10.1348/000709901158505

Van Droogenbroeck, F. V., and Spruyt, B. (2014). To stop or not to stop: an empirical assessment of the determinants of early retirement among active and retired senior teachers. *Res. Aging* 36, 753–777. doi: 10.1177/0164027513519449

Van Vegchel, N., De Jonge, J., Bosma, H., and Schaufeli, W. (2005). Reviewing the effort–reward imbalance model: drawing up the balance of 45 empirical studies. *Soc. Sci. Med.* 60, 1117–1131. doi: 10.1016/j.socscimed.2004.06.043

Varol, Y. Z., Weiher, G. M., Wendsche, J., and Lohmann-Haislah, A. (2021). Difficulties detaching psychologically from work among German teachers: prevalence, risk factors and health outcomes within a cross-sectional and national representative employee survey. *BMC Public Health* 21:2046. doi: 10.1186/s12889-021-12118-4

Wang, J. L., Lesage, A., Schmitz, N., and Drapeau, A. (2008). The relationship between work stress and mental disorders in men and women: findings from a population-based study. *J. Epidemiol. Community Health* 62, 42–47. doi: 10.1136/jech.2006.050591

Wang, Y., Ramos, A., Wu, H., Liu, L., Yang, X., Wang, J., et al. (2015). Relationship between occupational stress and burnout among Chinese teachers: a cross-sectional survey in Liaoning, China. *Int. Arch. Occup. Environ. Health* 88, 589–597. doi: 10.1007/s00420-014-0987-9

Wendsche, J., and Lohmann-Haislah, A. (2017). A meta-analysis on antecedents and outcomes of detachment from work. *Front. Psychol.* 7:02072. doi: 10.3389/fpsyg.2016.02072

Wepfer, A. G., Allen, T. D., Brauchli, R., Jenny, G. J., and Bauer, G. F. (2018). Work-life boundaries and well-being: does work-to-life integration impair well-being through lack of recovery? *J. Bus. Psychol.* 33, 727–740. doi: 10.1007/s10869-017-9520-y

Zurlo, M. C., Pes, D., and Siegrist, J. (2010). Validity and reliability of the effort-reward imbalance questionnaire in a sample of 673 Italian teachers. *Int. Arch. Occup. Environ. Health* 83, 665–674. doi: 10.1007/s00420-010-0512-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 February 2022
doi: 10.3389/fpsyg.2022.816358

To Those Who Have, More Will Be Given? Effects of an Instructional Time Reform on Gender Disparities in STEM Subjects, Stress, and Health

Nicolas Hübner[1], Wolfgang Wagner[2], Jennifer Meyer[3] and Helen M. G. Watt[4]*

[1] Institute of Education, University of Tübingen, Tübingen, Germany, [2] Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany, [3] Leibniz Institute for Science and Mathematics Education (IPN), Kiel, Germany, [4] School of Education and Social Work, The University of Sydney, Sydney, NSW, Australia

OPEN ACCESS

Edited by:
Joyce Endendijk,
Utrecht University, Netherlands

Reviewed by:
Maura Pilotti,
Prince Mohammad Bin Fahd University, Saudi Arabia
Johanna Sophie Quis,
Leibniz University Hannover, Germany
Simon Calmar Andersen,
Aarhus University, Denmark

****Correspondence:***
Nicolas Hübner
nicolas.huebner@uni-tuebingen.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *16 November 2021*
Accepted: *07 January 2022*
Published: *21 February 2022*

Citation:
Hübner N, Wagner W, Meyer J and Watt HMG (2022) To Those Who Have, More Will Be Given? Effects of an Instructional Time Reform on Gender Disparities in STEM Subjects, Stress, and Health. Front. Psychol. 13:816358. doi: 10.3389/fpsyg.2022.816358

Educational reformers all around the globe are continuously searching for ways to make schools more effective and efficient. In Germany, this movement has led to reforms that reduced overall school time of high track secondary schools from 9 to 8 years, which was compensated for by increasing average instruction time per week in lower secondary school (Grades 5–10). Based on prior research, we assumed that this reform might increase gender disparities in STEM-related outcomes, stress, and health because it required students to learn similar content in less amount of time. Therefore, we investigated how the school time reform affected gender disparities at the end of upper secondary school between 2011 and 2013. Specifically, we considered representative data of the last two cohorts who completed lower secondary school before the reform (N = 2,405) and the first two cohorts after the reform (N = 2,413) from the National Educational Panel Study. Potential differences in gender disparities were investigated for upper secondary school outcomes of subject-specific standardized test performance, self-concept, and interest in mathematics, biology and physics, as well as outcomes of school-related stress and health. Overall, we found substantial disparities between girls and boys, which seemed to change little after the reform. Exceptions were the statistically significant gender × reform interactions for one stress dimension (Overload) and two health dimensions (Overburdening and Achievement-related fear) which increased for both boys and girls, but more strongly for girls.

Keywords: school reform, instructional time, gender disparities, STEM, achievement, motivation, stress, health

INTRODUCTION

The optimal amount of time needed to learn has a longstanding history of research and critical socio-political discussions (Pischke, 2007; Cuban, 2008). As summarized by Patall et al. (2010), whereas proponents suggest that more instructional time (e.g., in a given school year) improves student achievement, opponents have called this into question. In their systematic review,

Patall et al. (2010) provided tentative evidence of the positive effects of increasing school time on student achievement, while reminding readers that much of what we currently know about this topic is based on weak designs. Considering further studies, findings on the effect of increasing instructional time on student achievement seem to be mixed, with some studies suggesting positive effects (e.g., Lavy, 2015; Andersen et al., 2016) and others finding zero or even negative effects (e.g., Allensworth et al., 2009; Nomi and Allensworth, 2009; Domina et al., 2015).

In contrast to these findings and intentions to increase instructional time, discussions regarding the optimal degree of time to learn went in a slightly different direction in Germany, where reforms from the past two decades were focused on making schools more efficient, for instance the "Gymnasium [high track secondary school] in 8 years"-reform (G8-reform). This reform aimed at reducing overall school time of high track secondary schools from 9 (G9) to 8 years (G8), which was compensated for by increasing average instruction time per week in lower secondary school (e.g., on average 3.69 additional hours per week each year; Homuth, 2017). Typically these reforms were implemented by either abolishing Grade 11 in upper secondary school or abolishing Grade 10 in lower secondary school (Kühn et al., 2013). Notably, in this study we focused on students from one German state (Baden-Württemberg). Here, overall instructional time per week was increased, while instruction time in STEM subjects remained largely comparable before and after the reform. Beyond this, further changes were implemented, which were required to increase instruction time per week, for instance new educational standards and school-specific curricula. Current research on the G8- reform is mixed in that some studies find student achievement to increase in lower secondary school (Huebener et al., 2017), whereas others find zero or negative effects on achievement, negative effects on stress levels and health, and delayed university enrollment of females (e.g., Büttner and Thomsen, 2015; Hübner et al., 2017a; Quis, 2018; Meyer et al., 2019; Marcus et al., 2020). Further studies are needed to investigate potential causes of these reported differences which may result from different samples (e.g., from different states) but also relate to the timeframe examined over which effects might accumulate or dissipate.

Although many school time studies focused STEM subjects, gender disparities, for instance on motivational outcomes or wellbeing, have been rarely investigated. This is surprising, because recent studies continue to find gender differences in STEM subjects (e.g., Watt, 2004; Else-Quest et al., 2010; Hübner et al., 2017b; Lazarides and Lauermann, 2019; Makarova et al., 2019; OECD, 2019) and on wellbeing (e.g., Hampel and Petermann, 2006; Moksnes et al., 2010; Salmela-Aro and Tynkkynen, 2012). Both motivation and wellbeing were found to be relevant for student achievement, aside from their importance in and of themselves (e.g., Widlund et al., 2018; Watt et al., 2019; Eccles and Wigfield, 2020; Wu et al., 2021). In addition, these constructs might also be affected by school reforms, as shown in prior studies (e.g., Hübner et al., 2017b; Marcus et al., 2020). It is consequently important to investigate whether girls may be disadvantaged relative to boys by the reform-induced changes, particularly regarding motivation and wellbeing. Therefore, in this study, we investigate gender disparities before and after the G8 school time reform in one German state (Baden-Württemberg) on an extended range of STEM-related outcomes beyond standardized test performance, such as subject-specific self-concept and interest in the subjects mathematics, biology and physics, and also include measures of school-related stress and health in the last year of secondary school.

GENDER AND SCHOOL TIME

Achievement, Gender, and School Time

Scarce evidence exists on gender disparities as a result of school time interventions or reforms. This is surprising for different reasons. First, gender equality is a central goal of all countries committed to human rights (United Nations General Assembly, 1948). Secondly, gender equality can contribute to economic growth (Altuzarra et al., 2021; Santos Silva and Klasen, 2021), particularly through increased participation in STEM jobs (Maceira, 2017; Hammond et al., 2020), which critically depend on achievement, self-concept, and course choices of STEM subjects in school (Updegraff et al., 1996; Parker et al., 2012; Watt et al., 2012, 2017; Schoon and Eccles, 2014. Referring to these arguments which underscore the relevance of monitoring effects of educational initiatives and reforms on gender disparities in general, it seems reasonable to believe that the G8-reform might specifically affect gender disparities in STEM. As girls and boys report different levels of self-concept and interest in math-intensive domains of STEM, which are central for subsequent achievement (e.g., Else-Quest et al., 2010; Hübner et al., 2017b, 2019; Eccles and Wigfield, 2020; Wu et al., 2021), it is important to investigate if the reform-induced intensifications/compression in lower secondary school might affect gender disparities in STEM-related achievement and motivation.

Several studies found differential effects of instructional time reforms for high- and low-performing students. For instance, Nomi and Allensworth (2009) investigated the effect of the "Double-Dose" algebra reform in Chicago, which required Grade 9 students with test scores below the national median to participate in additional algebra courses. The authors found a stronger positive effect for students close to the median, compared with students who performed much lower. In the same vein, Huebener et al. (2017) found small and sometimes non-significant changes in mathematics and science achievement for lower deciles of the performance distribution in the course of the G8-reform in Germany, whereas effects were larger for higher deciles. To our knowledge, that study is the only one in which the potential effects of the G8-reform on gender disparities were examined in science, reading and mathematics achievement for Grade 9 students. Interestingly, the findings suggested no statistically significant differential effects on girls and boys in Grade 9. The timing of assessment is important to consider when interpreting results of different G8-studies, because G8 students in Grade 9 have had substantially more instructional time compared with G9 students in Grade 9. However, by the

end of upper secondary school both cohorts have received a more comparable amount of instructional time.

In another study, Lavy (2015) reported that the treatment effect of increased school time was larger in higher performing countries, using PISA data. The author accounted for systematic differences between different countries by applying a country fixed-effects approach. These results provide tentative evidence of effect heterogeneity as a result of school time reforms, depending on students' level of achievement.

Many school time studies and reforms focused on changes in STEM achievement of high and low performers, while gender disparities, for instance on motivational outcomes or wellbeing, have been rarely investigated. This constitutes an important limitation of many prior studies because gender disparities in STEM are well documented: The OECD (2019) reported a mathematics advantage for boys in 32 economies/countries (of 78; 14 economies/countries reported advantages for girls) and a science advantage for girls in 34 countries (of 78; 9 economies/countries reported advantages for boys). Notably, the differences were small on average (d = 0.05; ranging from d = 0.22 in Colombia to a non-significant difference of d = 0.01 in the Netherlands), and recent research suggests closings of these gaps, for instance in science achievement (e.g., Meinck and Brese, 2020). There are also meta-analyses that essentially found very small gender differences in math achievement but substantial variability across countries (e.g., Else-Quest et al., 2010). However, robust and systematic gender differences favor boys for math self-concept and interest in adolescence (e.g., Watt, 2004; Else-Quest et al., 2010; Frenzel et al., 2010; Nagy et al., 2010; Hübner et al., 2017b, 2019; Widlund et al., 2018; Parker et al., 2020; Mejía-Rodríguez et al., 2021; Wu et al., 2021).

Probably most important in the context of this study, prior research using rich data from the end of German upper secondary school has provided evidence for substantial differences between boys and girls on a broad variety of mathematically intensive STEM outcomes, even after controlling for cognitive abilities. For instance, Hübner et al. (2019) found girls to have statistically significantly lower achievement in mathematics ($d \geq 0.45$, $p < 0.05$) and physics ($d \geq 0.63$, $p < 0.05$), compared to boys, whereas no such gender differences were found in biology. In addition, differences in mathematics in advantage of boys seem to be pronounced in Germany already by Grade 4 in elementary school (d = 0.18; Stanat et al., 2017).

Self-Concept, Interest, Gender, and School Time

Women and men differ substantially in regard to their mathematical and mathematics-intensive STEM educational pathways and career aspirations (Watt et al., 2012, 2017; Schoon and Eccles, 2014; Lazarides and Lauermann, 2019; Makarova et al., 2019; Lazarides et al., 2020). This process has been referred to as the leaky STEM pipeline (Jacobs and Simpkins, 2005). Prior research has found that central to the choice of advanced course enrollments are students' subject-specific achievement (Updegraff et al., 1996; Parker et al., 2012) and self-concept and values (Watt et al., 2012), even after controlling for prior achievement levels in the domain (Watt et al., 2017). These motivational variables have been linked not only to school enrollment but further to aspired educational and occupational pathways in mathematics and STEM subfields (Watt et al., 2012, 2017). Choosing advanced courses in high school constitutes a key factor for subsequent enrollment in STEM subjects at university (Ma and Johnson, 2008; Eccles and Wigfield, 2020; Lazarides et al., 2020). Thus, if a reform has differential effects on girls and boys (e.g., increases or decreases to their motivation), it is likely to affect subsequent decisions for or against related courses in high school or later on at university (e.g., Hübner et al., 2017b; Biewen and Schwerter, 2021).

This line of argumentation can be extended and implications can be derived more theoretically: Expectancy-value theory (Eccles, 1983; Eccles and Wigfield, 2002, 2020) outlines that key elements for choices are students' expectations of success and task values and that both are influenced by prior achievement. Empirical evidence for this assumption can be found, for instance, in literature on the reciprocal effects model between self-concept and achievement (Marsh and Craven, 2006; Seaton et al., 2015). Self-concept is defined as students' perceptions about their abilities, which develops via engagement with others (Shavelson et al., 1976; Marsh, 1990; Marsh et al., 2016). Task values, the other important set of variables to explain choices, consist of four components: intrinsic, attainment, utility, and cost values. Intrinsic value refers to students' enjoyment when performing a specific task, attainment value refers to the personal importance a student attaches to a task, and utility value refers to its usefulness; researchers have combined attainment and utility values and referred to "importance value." Costs, on the other hand, refer to the perceived negative consequences of task engagement, for example, effort or psychological and social costs (Watt et al., 2019).

Regarding subject-specific self-concept and interest, prior research suggests differences between girls and boys, which typically follow stereotypic patterns: Boys tend to report higher self-concept and interest in math-intensive STEM subjects compared to girls, whereas these effects are typically zero or in favor of girls in subjects such as biology (e.g., Denissen et al., 2007; Hübner et al., 2017b, 2019; Watt et al., 2017, 2019; Parker et al., 2020; Mejía-Rodríguez et al., 2021). Therefore, if school time reforms force girls to learn similar content in less amount of overall time in subjects they are less interested in and in which they have lower perceptions of their own abilities (e.g., girls in math-intensive STEM subjects), this might even reinforce such less positive perceptions (e.g., Hübner et al., 2017b, 2019). In addition, if the reforms differentially affect boys' and girls' STEM achievement this might also foster further disparities, for instance regarding students' self-concept, as these variables are reciprocally related (e.g., Marsh and Craven, 2006; Arens et al., 2017; Wu et al., 2021).

School-Related Stress, Health, Gender, and School Time

Other variables that are important to consider in the context of an intensified learning environment include students' perceived

stress and health. These variables might be particularly relevant in the context of increasing instructional time because it is intended that students spend more time with learning in school, which might lead to reduced or even too little leisure time to recover (Milde-Busch et al., 2010; Hübner et al., 2017a). As outlined in prior research, mental health is also associated with student achievement (e.g., Tuominen-Soini and Salmela-Aro, 2014; Fiorilli et al., 2017). For instance, Agnafors et al. (2021) found that students with mental health problems in very early years more often performed below grade level later on. Another study by Fiorilli et al. (2017) suggests that students' burnout is highly relevant for student achievement, both directly and indirectly. The importance of considering wellbeing as a foundation for students' aspirations was underscored in a study by Widlund et al. (2018) of Finnish students. Depending on the age group, the authors were able to identify either three (Grade 7) or four (Grade 9) latent profiles based on students' attainment and self-concept in mathematics, their engagement, and three burnout subscales. They found that students with negative academic wellbeing had statistically significantly lower aspirations compared to thriving students. Interestingly, they found that girls were overrepresented in the negative academic wellbeing profile, which is in line with prior findings on gender disparities in school burnout (Salmela-Aro and Tynkkynen, 2012).

Beyond these studies, further research has produced evidence suggesting that girls generally do develop higher stress levels, compared with boys (e.g., Hampel and Petermann, 2006; Moksnes et al., 2010). Studies that have focused on investigating school stressors found schoolwork pressure to partly explain psychological complaints and psychosomatic pain (Hjern et al., 2008), and girls reported higher levels of performance-related stress at school (Moksnes et al., 2010). Finally, prior research provides evidence that increasing learning time might lead to more stress-related health problems (e.g., Marcus et al., 2020). Related to this, Quis (2018) investigated gender-specific differences between G8- and G9-students on school-related stress and health among students at the end of upper secondary school. She found considerable differences in school-related stress and mental health before and after the reform, mainly driven by girls (health) or boys and girls (stress). However, uncertainty exists whether such effects result from increases on a majority of stress facets (e.g., feelings of exhaustion, achievement-related overburdening, or not being able to recover in leisure time), or particularly on specific facets and not others.

THE PRESENT STUDY

Based on our theoretical and empirical considerations above, three potential effects of the G8-reform on existing gender disparities in math-intensive STEM subjects can be derived, displayed in **Figure 1**. First, the *"perpetuation" model* would suggest no changes in disparities between boys and girls before and after the reform. This result pattern might be found, for instance, if the reform affected gender disparities in lower secondary school, where it was implemented, but these effects "washed out" by the end of upper secondary school, or if the reform-induced changes were too weak or equally affected boys and girls. Second, the "*accumulated advantages and disadvantages" model* would imply findings in the shape of the Matthew effect. This effect was first found by Merton (1968) and subsequently used by many researchers in educational, psychological and social scientific research to describe increasing disparities over time (e.g., for different ethnicities or students with different socio-economic backgrounds; e.g., Baumert et al., 2012). In the case of gender disparities, this effect would suggest that gender-specific advantages might increase (e.g., boys' advantages over girls on achievement, self-concept, and task values in mathematics and physics), leading to overall widened disparities. Finally, the "*compensation" model* would imply that the disadvantaged group improves more over time, leading to smaller disparities after the reform. This effect would be found if girls benefit more from the reform, for instance because additional time is used to practice curricular content rather than to learn additional content (e.g., Hübner et al., 2017b).

Most of the cited literature above provides evidence for the accumulated advantages and disadvantages model, whereby school time reforms might particularly benefit higher performing students (Nomi and Allensworth, 2009; Lavy, 2015; Huebener et al., 2017), which would, in our case, imply widening gender-specific disparities on math-intensive STEM outcomes. Regarding STEM subjects, it is also important to consider hours per week in G8 vs. G9. Doing this, we found minor differences in officially reported hours in lower secondary school. Despite this, prior studies reported differences in student achievement between G8 and G9 students (Huebener et al., 2017; Hübner et al., 2017a). In our view, these findings underscore that it is important to not only consider subject-specific instructional time in school, but time spent on school-related purposes as a whole (e.g., Scheerens, 2014). For instance, even if instructional time were comparable in STEM subjects in G8 and G9, the overall instructional time per week in lower secondary school in G8 increased, which had an impact on the amount of time at home and students' leisure time (Milde-Busch et al., 2010; Hübner et al., 2017a). Time at home constitutes a quite important predictor for school performance, for instance because students' school-related engagements with parents can contribute to their learning (Berkowitz et al., 2015), investing time in homework might improve student achievement (Rawson et al., 2017), and leisure time can be used for addressing specific learning gaps, preparing for exams outside from school, or to recover from school-related stress (Milde-Busch et al., 2010). Further, girls were found to invest more time at home for school-related purposes (Wagner et al., 2008), which might also explain potential differential effects of the G8-reform. From this perspective, even if instructional time in STEM subjects remains comparable, if students have to invest more time, on average, in formal schooling and have less time for self-paced learning, learning activities at home, or relaxation, this might have detrimental effects on their achievement and wellbeing.

As outlined above, disparities between boys and girls were inconsistent and small at most in mathematics achievement, substantially larger in math self-concept and non-existent in science (Watt, 2004; Else-Quest et al., 2010; Watt et al., 2012;

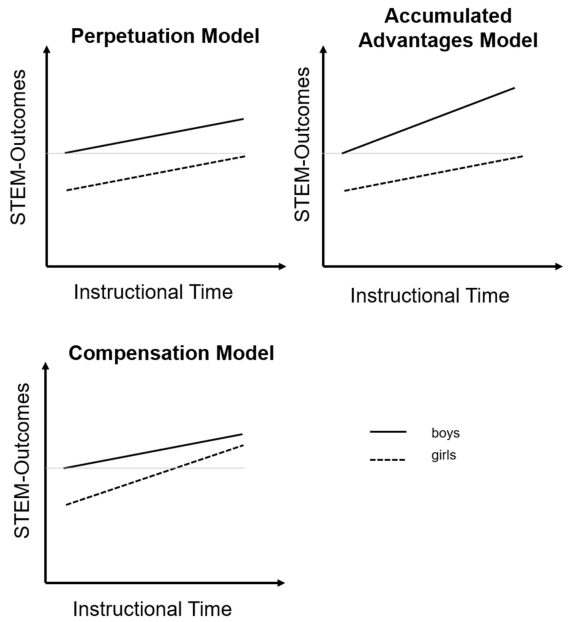


FIGURE 1 | Hypothetical effects of the school time reform on gender disparities in math-intensive STEM subjects.

OECD, 2019). Based on this, we expect zero or very small effects on math or math-intensive STEM achievement, larger effects on math-intensive STEM self-concept and interest, but null effects for biology. Regarding stress and health, it seems reasonable to believe that the reform might be perceived as more demanding by girls compared to boys, which might produce larger differences between boys and girls after the reform. As girls report higher levels of burnout and stress than boys (e.g., Salmela-Aro and Tynkkynen, 2012; Widlund et al., 2018), increasing demands of the learning environment might particularly be harmful for them. Prior research has (on average) found larger disparities between boys' and girls' school-related stress levels after the reform (Quis, 2018), but has not yet explored whether average differences might mask differences on specific stress facets but not others. We will extend findings based on unidimensional models to obtain a nuanced understanding of gender-specific reform effects on different dimensions of stress and health.

MATERIALS AND METHODS

Description of the Study and Sample

We used data from the Additional Study Baden-Württemberg (Blossfeld et al., 2011) from the National Educational Panel Study (NEPS; Scientific Use File 3.2.0). The dataset contains representative data for Baden-Württemberg, assessed from four different cohorts in the final semester of upper secondary school. Two cohorts completed German lower secondary school before the reform and two completed it after the reform. We compared outcomes of these cohorts assessed at the end of upper secondary school (G9: Grade 13 or G8: Grade 12). This design is typically referred to as a cohort control design (Shadish et al., 2002). Overall, students from 44 high track upper secondary schools participated in the study: Cohort 1 (before the reform): n = 1,226 (55% girls); Cohort 2 (before the reform): n = 1,179 (55% girls); Cohort 3 (after the reform): n = 1,205 (56% girls); Cohort 4 (after the reform): n = 1,208 (55% girls). Before the reform, students graduated after 9 years of high track upper secondary school, whereas after the reform students graduated after 8 years. The first cohort of students graduated in 2011, the second (Grade 13) and third (Grade 12) in 2012, and the fourth in 2013. Notably, in Baden-Württemberg, Grade 11 was abolished to implement the G8-reform (Kühn et al., 2013). Data were collected in the final semester of the last year of upper secondary school. Students in Germany are required to spend at least 265 h per week each year in school. This means that G9 students are required to spend on average 265/9 = 29.44 h per week each year in school, whereas G8 students are required to spend 265/8 = 33.13 years per week each year in school, reflecting a difference of 3.69 additional hours that students in G8 are required to spend per week in school. Overall, cumulated mandatory hours were 11 h higher for G8 students from grade 5 to grade 6, and 16 h higher for G8 students from grade 7 to grade 10 in Baden-Württemberg (Homuth, 2017).

Instruments

In all cohorts, identical instruments were administered to assess subject-specific standardized achievement, self-concept, and interest in the subjects mathematics, biology and physics, as well as to assess school-related stress and health. The questionnaire is available in the NEPS data center[1].

Standardized Test Performance

Comprehensive information on these tests and different quality indicators can be found in the scaling reports of the National Educational Panel Study (Duchhardt, 2015; Hübner et al., 2016a,b). The mathematics test was based on 20 items from the four areas of quantity, space/shape, change/relationships, and data/chance (Duchhardt, 2015). The biology test consisted of 60 items from the areas of cytology/anatomy/metabolism, information processing/characteristics/immunology, genetics/development biology, ecology, and systematics/evolution (Hübner et al., 2016a). Finally, physics achievement was assessed using 41 items from nine different areas, for instance electrical fields and interdependency, waves, and optics (Hübner et al., 2016b). In our sample, the reliability of the weighted likelihood estimator (WLE; Adams, 2005) was *Rel.* = 0.70 for the math test, *Rel.* = 0.61 for the physics test, and *Rel.* = 0.73 for the biology test. As outlined below, latent variable models were specified to adequately address their measurement error.

Subject-Specific Self-Concept

Subject-specific self-concept was assessed using four items from the translated Self-Description Questionnaire III (Marsh and O'Neill, 1984) for each of the subjects mathematics,

[1]https://www.neps-data.de/Data-Center/Data-and-Documentation

biology and physics. For example, students were asked to rate their agreement to: "I have never done well in mathematics" or "I am good at mathematics" on a 4-point rating scale from 1 (*does not apply at all*) to 4 (*completely applies*). Negatively formulated items were reverse coded. Cronbach's α for students' self-concept was α = 0.94 for mathematics, α = 0.91 for biology, and α = 0.94 for physics.

Subject-Specific Interest

Subject-specific interest was assessed using four items based on the expectancy-value framework (Eccles, 1983; Eccles and Wigfield, 2002) for each of the subjects mathematics, biology, and physics. Items were comparable to those from prior German large-scale studies (Trautwein et al., 2006, 2010). For instance, students were asked to rate their agreement to: "It is important for me personally to be good at mathematics" or "Math is just exciting for me" on a 4-point rating scale from 1 (*does not apply at all*) to 4 (*completely applies*). Negatively formulated items were reverse coded. Cronbach's α for students' interest was α = 0.82 for mathematics, α = 0.87 for biology, and α = 0.90 for physics.

School-Related Stress

School-related stress was assessed using 15 items (Hübner et al., 2017a). Example items are: "Sometimes I have trouble falling asleep because problems from school are on my mind," "Even during my free time I think about troubles at school," or "Pressure at school is too high" (see **Supplementary Table 1** for a full list of items). Students were asked to answer these items on a 4-point rating scale ranging from 1 (*completely disagree*) to 4 (*completely agree*). The stress scale constitutes an instrument which was developed by the NEPS (including internal review cycles), which has a specific focus on school-related stress. Both instruments were also administered in the NEPS Thuringia study (Blossfeld et al., 2011). Negatively formulated items were reverse coded. Reliability of the scale was high (Cronbach's α = 0.91).

Health

Students' health was measured by asking them to rate how often they experienced 26 different health problems on a rating scale from 1 (*never*) to 4 (*more than 6 times during the last 6 weeks*), respectively (Bergmüller, 2007). Among others, health problems such as "headaches," "sleep disturbances," "vomiting," or "feelings of inner emptiness" were assessed (see **Supplementary Table 2** for a full list of items). There are further studies, which administered comparable health items, particularly in the field of medical science (e.g., Milde-Busch et al., 2010), but also beyond (Bergmüller, 2007). The health scale was administered in prior cycles of the PISA study (Bergmüller, 2003). Reliability of the scale was high (Cronbach's α = 0.92).

In examining these outcomes, we controlled for a variety of covariates in the adjusted models. These were immigration background (i.e., students with at least one parent born abroad), number of available books at home, highest international socioeconomic index in the family (HISEI), non-verbal cognitive skills (i.e., perceptual speed and reasoning; Haberkorn and Pohl, 2013), and whether students had repeated a class. In addition, we controlled for the course level (advanced, basic, or de-selection) when investigating differential effects on standardized test performance. An overview on course enrollment by gender and subject is given in **Supplementary Table 5**. Notably, there were no gender differences in math enrollment, as all students are mandated by law to enroll in advanced mathematics courses (4 h per week), whereas differences were most visible in physics, where only 8.7% of girls were enrolled in advanced courses, compared to 29.9% of boys.

Statistical Analysis

The main analysis proceeded in two steps. First, we estimated multiple-group models for the eight different groups (4 cohorts × gender) in M*plus* 8.6 (Muthén and Muthén, 1998–2017). We did this separately for standardized test achievement, self-concept, interest, school-related stress, and health. For achievement, we used multidimensional (multiple-group) item response theory (IRT) models (see Jöreskog and Goldberger, 1975; Hübner et al., 2020). For the remaining constructs traditional structural equation models (SEMs) were applied. Prior work offered clear guidance on how to define measurement models for the achievement measures, self-concept, and interest (e.g., Marsh, 1992; Eccles and Wigfield, 2002; Duchhardt, 2015; Hübner et al., 2016a); this was not the case for the instruments used to assess students' stress and health which were typically analyzed as a single aggregate score and not with multidimensional models (e.g., Bergmüller, 2003; Hübner et al., 2017a; Quis, 2018). In this study, we utilized a data-driven procedure to explore the underlying factor structure of stress and health items using exploratory structural equation models (ESEMs). As outlined by Marsh et al. (2014), ESEMs combine useful features of exploratory and confirmatory factor analysis (EFA/CFA) such as confirmatory tests of factor structures and associations between different latent factors, and they allow small cross-loadings. For school-related stress and health we performed ESEMs with geomin rotated factor loadings in a multiple group framework. To identify the most adequate solution, we first specified different (single-group) ESEM models with an increasing number of latent factors, before running ESEMs in a multiple group framework with the eight groups (gender × cohort). Models were constrained to test strong factorial/scalar measurement invariance, which is required to meaningfully compare latent means across groups. To judge model fit, we considered the Comparative Fit index (CFI), the Tucker-Lewis index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). Based on prior research (MacCallum et al., 1996; Hu and Bentler, 1999; Yu, 2002), we considered the following cutoffs to indicate good model fit: CFI and TLI $\geq$ 0.95, SRMR and RMSEA $\leq$ 0.05.

Using these models, we compared the means or—for models with covariates—intercepts of the latent outcomes between the resulting groups using the delta method (Oehlert, 1992) by applying the MODEL CONSTRAINT option in M*plus*.

Statistically significant differences between the specific group differences constitute interaction effects. We estimated (a) gender differences in G9 cohorts and (b) gender differences in G8 cohorts, and one interaction effect: (c) the difference between a and b (reform × gender). We specified unadjusted models without covariates and adjusted models including covariates (e.g., cognitive abilities, socioeconomic background; see "Instrument" section) to check the robustness of our results. To better interpret our findings, results were transformed into a metric with an overall $M = 500$ and $SD = 100$ for achievement and to a metric with an overall $M = 50$ and $SD = 10$ for the remaining constructs, using the pooled variance of the latent variables from the unadjusted models. For consistency, we report two-sided p-values throughout, although prior studies suggest a directional hypothesis for stress, health, and math-intensive STEM self-concept in disadvantage of girls. We therefore interpret one-sided p-values to judge statistical significance for those constructs (one-sided p-value = two-sided p-value/2). For all other outcomes, no consistent directional hypothesis could be derived from the literature. All models were specified using full information maximum likelihood estimation (FIML; Enders, 2010), robust standard errors (McNeish et al., 2017), and survey weights.

RESULTS

Preliminary Analysis

First, we inspected descriptive statistics. As shown in **Table 1**, overall, differences between the two cohorts were small. Only with regard to perceptual speed, students in G9 scored slightly higher. Further, students in G8 repeated classes slightly less often than students in G9. This resulted from a generally low repetition rate due to a specific feature of the reform implementation: If students from the last G9 cohort were required to repeat a grade, they had to move from, for instance, the end of grade 10 to the beginning of grade 9, because the respective grade 10 cohort in G8 would have already been ahead of the grade 10 in G9, which the student should repeat (due to the additional hours per week in lower secondary school). These differences were controlled for in the adjusted models as outlined below.

Gender-Specific Differences Before and After the Reform

Next, we inspected gender-specific differences. As visible from **Table 2**, we found substantial differences between girls and boys, both before and after the reform.

Standardized Test Performance

Regarding standardized test performance, boys were found to score statistically significantly higher than girls before the reform in biology ($b = 18.84$, $p = 0.001$), in mathematics ($b = 79.23$, $p < 0.001$), and in physics ($b = 91.85$, $p < 0.001$). Differences were smaller in biology and substantially larger in mathematics and physics, and these differences remained equally pronounced after the reform. After the reform, the respective differences amounted to $b = 18.61$ ($p = 0.003$) in biology, $b = 70.60$ in mathematics ($p < 0.001$), and $b = 92.64$ points ($p < 0.001$) in physics. Notably, differences between gender disparities from before and after the reform (i.e., the gender × reform interaction effect) were not statistically significant for any standardized test performance. This coefficient amounted to $\Delta b = 0.23$ points ($p = 0.975$) in biology, $\Delta b = 8.63$ ($p = 0.254$) in mathematics, and $\Delta b = -0.79$ ($p = 0.915$) in physics. These results suggest that differences between girls and boys were generally large on these standardized test outcomes before the reform and remained comparably large after the reform, consistent with the perpetuation model (see **Figure 1**).

Subject-Specific Self-Concept

With regard to subject-specific self-concept, we found a slightly different picture. Here, no statistically significant differences between boys and girls were found for biology, before ($b = 0.00$, $p = 0.999$) or after the reform ($b = -0.07$, $p = 0.870$). Regarding mathematics, girls and boys differed statistically significantly before the reform ($b = 4.06$, $p < 0.001$) and after the reform ($b = 4.82$, $p < 0.001$), with boys having higher self-concept scores. The differences in gender disparities before vs. after the reform

TABLE 1 | Descriptive statistics on central covariates before and after the reform.

	G9		G8			
	$n = 2.405$		$n = 2.413$			
Variable	*M*	*SD*	*M*	*SD*	*ES*	*p*
Immigration background (1 = yes)	0.23	0.42	0.22	0.41	1%	0.349
Books at home	4.72	1.24	4.73	1.25	−0.01	0.858
HISEI	58.16	15.42	58.41	15.50	−0.02	0.653
Perceptual speed	65.32	11.41	64.98	11.96	0.03	0.660
Reasoning	10.80	1.26	10.71	1.27	0.07	0.023
Class repeater (1 = yes)	0.10	0.30	0.06	0.24	4%	< 0.001

Descriptive statistics were estimated using full information maximum likelihood estimation, cluster-robust standard errors, and survey weights. HISEI = highest international socioeconomic index in the family. ES = Effect size. We used Cohen's d (Cohen, 1988) for continuous variables, which was estimated as M_{G9}-M_{G8} divided by the pooled SD, and differences in percentage points for dichotomous variables. Please also see Hübner et al. (2017a) and Quis (2018) for additional tests of potential selectivity and representativeness and comparisons of differences on covariates across different cohorts. Additional information on the estimation of the survey weights can be found in Schönberger and Aßmann (2014).

TABLE 2 | Unadjusted gender-disparities before and after the reform on standardized test performance, subject-specific self-concept and interest, and school-related stress and health.

	b_{G9}	SE	*p*	b_{G8}	SE	*p*	Δ *b*	SE	*p*
Standardized test performance									
Biology	**18.84**	5.80	0.001	**18.61**	6.10	0.003	0.23	8.85	0.975
Mathematics	**79.23**	5.75	<0.001	**70.60**	5.31	<0.001	8.63	7.64	0.254
Physics	**91.85**	6.05	<0.001	**92.64**	5.19	<0.001	−0.79	6.92	0.915
Subject-specific self-concept									
Biology	0.00	0.54	0.999	−0.07	0.42	0.870	0.07	0.67	0.917
Mathematics	**4.06**	0.48	<0.001	**4.82**	0.46	<0.001	−0.76	0.64	0.231
Physics	**6.55**	0.48	<0.001	**7.02**	0.48	<0.001	−0.47	0.68	0.488
Subject-specific interest									
Biology	−0.59	0.59	0.320	−1.14	0.53	0.030	0.55	0.72	0.446
Mathematics	**1.89**	0.56	0.001	**2.76**	0.50	<0.001	−0.87	0.80	0.268
Physics	**6.48**	0.54	<0.001	**5.53**	0.64	<0.001	0.95	0.82	0.238
School-related stress									
Difficulties to relax	**−7.34**	0.69	<0.001	**−8.58**	0.71	<0.001	1.24	0.86	0.149
Exhaustion	**−5.03**	0.63	<0.001	**−5.60**	0.53	<0.001	0.57	0.72	0.422
Overload	**−2.00**	0.51	<0.001	**−3.61**	0.75	<0.001	**1.61**	0.73	0.027
Malaise	**2.60**	0.93	0.005	1.49	1.09	0.171	1.11	0.75	0.134
Alignment issues	**−3.15**	0.67	<0.001	**−4.00**	0.77	<0.001	0.85	0.75	0.256
Health									
Overburdening	**−3.69**	0.74	<0.001	**−5.01**	0.70	<0.001	**1.32**	0.74	0.076
Achievement-related fear	**−6.12**	0.72	<0.001	**−7.81**	0.70	<0.001	**1.69**	0.64	0.008
Diverse symptoms	**−13.05**	2.56	<0.001	**−14.08**	2.52	<0.001	1.03	1.00	0.301
Uneasiness	−1.77	2.11	0.352	−2.25	1.90	0.285	0.49	0.69	0.476
Depressive symptoms	**−3.60**	0.66	<0.001	**−4.63**	0.73	<0.001	1.02	0.65	0.119
Gastrointestinal issues	**−3.62**	1.48	0.004	**−4.01**	1.11	0.012	0.39	0.72	0.492

b_{G9} = Gender differences before the reform; b_{G8} = Gender differences after the reform. Positive values indicate higher values for boys. Δb = Difference of gender differences before (G9) minus after (G8) the reform. The metric of the latent variable was transformed to M = 500 and SD = 100 for standardized test performance and to M = 50 and SD = 10 for all other outcomes using pooled means and standard deviations. Statistically significant differences ($p < 0.05$) are printed in bold. Regresson coefficients (b's) are based on group mean differences in a multiple group model. Two-sided p-values are reported. In cases where we had a directional hypothesis based on prior literature (e.g., higher stress scores of girls), one-sided p-values should be calculated/interpreted, which can be calculated by dividing the reported two-sided p-value by 2.

did not reach statistical significance ($\Delta b = -0.76$, $p = 0.231$). Finally, regarding physics, a similar picture as in mathematics emerged. Boys had higher scores before ($b = 6.55$, $p < 0.001$) and after ($b = 7.02$, $p < 0.001$) the reform, and these differences did not change ($\Delta b = -0.47$, $p = 0.488$).

Subject-Specific Interest

Next, we had a closer look at the results for subject-specific interest. The results were fairly similar to those for subject-specific self-concept, however, gender differences were less pronounced in mathematics. Here differences amounted to 1.89 points ($p = 0.001$) before the reform and 2.76 points after the reform ($p < 0.001$). The reform × gender interaction effect did not reach statistical significance ($\Delta b = -0.87$, $p = 0.268$). In summary, the results for achievement test performance, subject-specific self-concept, and subject-specific interest provided evidence in support of the perpetuation model.

School-Related Stress

Subsequently, we investigated potential differences for school-related stress. To do this, we first fitted a series of ESEM models with an increasing number of latent factors. The solution to first reach adequate model fit (CFI and TLI ≥ 0.95 and RMSEA and SRMR ≤ 0.05) was a model with six factors; however, one factor had substantial loadings only on the (reverse-coded) negatively worded items, while the loadings of these items on all other factors were small (all ≤ 0.06 for t5m and ≤ 0.02 for t5n). Also considering findings from prior studies on challenges of considering negatively worded items of instruments (e.g., DiStefano and Motl, 2006; van Sonderen et al., 2013; Zhang et al., 2016), we decided to drop the two reverse-scored items, which resulted in a more parsimonious five-factor multiple group model [unadjusted model: $\chi^2(520) = 833.378$, $p < 0.001$, CFI = 0.99, TLI = 0.98, RMSEA = 0.03, SRMR = 0.03]. From a substantive perspective, this model was comparable to the model with six factors but did not include the factor for the negatively worded items. As a robustness check, we also specified a model in which we predicted the previously dropped (reverse coded) items t5n and t5m by the five factors, a reform dummy variable, gender, and the interaction term reform × gender. Our findings showed that, after conditioning on the five factors, none of the remaining variables was statistically significantly associated with the t5n or t5m variable. Therefore, it seems unlikely that dropping the two negatively worded items had

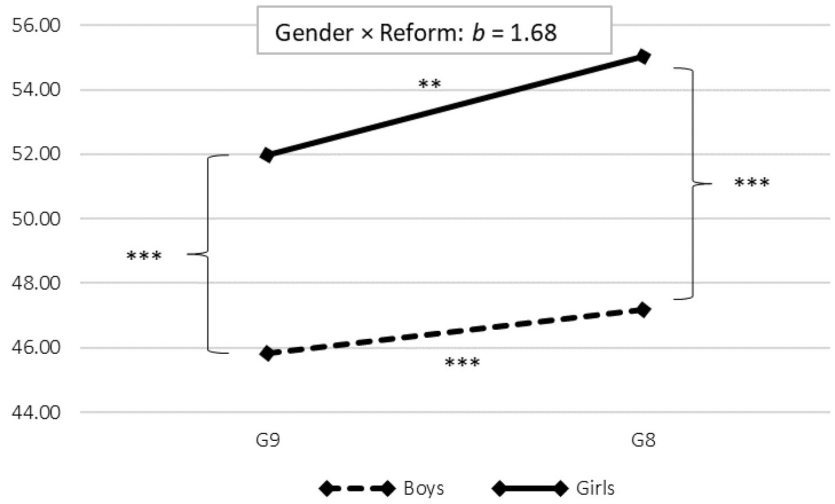


FIGURE 2 | Gender-specific interaction effect for achievement-related fear. Based on findings reported in **Table 3**. ***$p < 0.001$. **$p < 0.01$.

a substantial impact on our main research question. The five factors were given names based on their loading patterns (see **Supplementary Table 3**): (1) Difficulties to relax, (2) Exhaustion, (3) Overload, (4) Malaise, and (5) Alignment issues. As is visible in **Table 2**, we found statistically significant differences between boys and girls on all factors in G9 (all $ps \leq 0.005$) and on all factors besides Malaise ($p = 0.171$) in G8 (all $ps < 0.001$). Whereas these differences generally suggested higher stress levels for girls on four of five factors (Difficulties to relax, Exhaustion, Overload, and Alignment issues), boys in G9 reported having more issues on the Malaise factor. Finally, we found a statistically significant gender × reform interaction effect on the Overload factor ($\Delta b = 1.61$, $p = 0.027$). This factor had its highest loadings on items such as "Pressure at school is too high" or "I consider the requirements at school in general as stressful." The interaction effect indicated that the difference between boys and girls on this factor was larger in G8 than G9. Further explorations revealed that it was strongly driven by larger overload stress levels for girls in G8 vs. G9 ($\Delta b = 5.21$, $p < 0.001$), compared to boys ($\Delta b = 3.60$, $p < 0.001$).

Health

For health, we found an ESEM model with six factors to reach the cutoff values for model fit as outlined above [unadjusted model: $\chi^2(2,452) = 4{,}041.463$, $p < 0.001$, CFI = 0.95, TLI = 0.95, RMSEA = 0.03, SRMR = 0.04]. The six factors were given names based on their loading patterns (see **Supplementary Table 4**): (1) Overburdening, (2) Achievement-related fear, (3) Diverse symptoms, (4) Uneasiness, (5) Depressive symptoms, and (6) Gastrointestinal issues. The results pointed in the same direction as for stress: Girls tended to have statistically significantly more health issues on all six health factors, although the difference on the Uneasiness factor between boys and girls in G9 and G8 cohorts was not statistically significant (see **Table 2**). The largest difference was found on the Diverse symptoms factor, which had as its three highest loadings the indicators "Headaches," "Bad dreams," and "Stomach ache" (G9: $b = -13.05$, $p < 0.001$; G8: $b = -14.08$, $p < 0.001$).

For health, we found two statistically significant gender × reform interaction effects on the factors Overburdening (highest loadings for "Difficulty concentrating," "Tiredness, fatigue," and "Easily irritable") and Achievement-related fear ("Feeling that excessive demands are being made of me," "Fear of going to school," "Fear that it's all getting too much"). For Overburdening, this interaction effect amounted to $\Delta b = 1.32$ ($p = 0.076$ [$p_{one-sided} = 0.038$]), whereas for Achievement-related fear, it amounted to $\Delta b = 1.69$ ($p = 0.008$). The interaction effect for Achievement-related fear is displayed in **Figure 2**, which increased more for girls than boys following the reform.

Finally, we compared results from the unadjusted and adjusted models (see **Table 3**), in which we controlled for further covariates such as cognitive abilities and socioeconomic background. Overall, we did not find substantial differences between the two solutions, in terms of statistical significance or the direction or size of coefficients (see **Tables 2**, **3**). Our results for achievement, self-concept, and interest provide tentative evidence in line with the perpetuation model, whereas our findings for stress and health are more in line with the accumulated advantages/disadvantages model (see **Figure 1**).

DISCUSSION

In this study, we investigated the effects of the G8-reform on gender disparities in STEM achievement, self-concept, and interest, as well as school-related stress and health. To do this, we compared data of four successive student cohorts, two from before the reform and two from afterward. Specifically, the reform changed the overall school time of high track secondary schools from 9 to 8 years, which was compensated for by increasing average instruction time per week in lower secondary school (Grades 5–10 in Germany).

TABLE 3 | Adjusted gender-disparities before and after the reform on standardized test performance, subject-specific self-concept and interest, and school-related stress and health.

	b_{G9}	SE	*p*	b_{G8}	SE	*p*	Δ *b*	SE	*p*
Standardized test performance									
Biology	**19.29**	6.41	0.002	**19.98**	5.80	0.001	−0.69	8.85	0.943
Mathematics	**78.57**	5.53	<0.001	**68.44**	4.76	<0.001	10.13	6.86	0.142
Physics	**89.90**	4.61	<0.001	**90.77**	5.19	<0.001	−0.86	5.62	0.872
Subject-specific self-concept									
Biology	−0.01	0.54	0.992	−0.19	0.42	0.650	0.18	0.68	0.783
Mathematics	**4.02**	0.44	<0.001	**4.45**	0.42	<0.001	−0.42	0.59	0.468
Physics	**6.55**	0.43	<0.001	**7.01**	0.44	<0.001	−0.46	0.62	0.467
Subject-specific interest									
Biology	−0.51	0.57	0.369	−0.96	0.56	0.088	0.45	0.74	0.550
Mathematics	**1.94**	0.53	<0.001	**2.37**	0.50	<0.001	−0.43	0.75	0.564
Physics	**6.55**	0.51	<0.001	**5.43**	0.59	<0.001	1.12	0.78	0.147
School-related stress									
Difficulties to relax	**−7.42**	0.71	<0.001	**−8.40**	0.81	<0.001	0.98	0.90	0.271
Exhaustion	**−5.08**	0.66	<0.001	**−6.09**	0.49	<0.001	1.01	0.72	0.161
Overload	**−2.05**	0.51	<0.001	**−3.68**	0.70	<0.001	**1.63**	0.69	0.017
Malaise	**2.82**	0.97	0.004	1.48	1.10	0.176	1.34	0.87	0.123
Alignment issues	**−3.29**	0.67	<0.001	**−4.14**	0.82	<0.001	0.86	0.77	0.265
Health									
Overburdening	**−3.63**	0.70	<0.001	**−4.90**	0.74	<0.001	**1.28**	0.75	0.089
Achievement-related fear	**−6.17**	0.69	<0.001	**−7.85**	0.72	<0.001	**1.68**	0.65	0.010
Diverse symptoms	**−12.36**	2.66	<0.001	**−13.31**	2.84	<0.001	0.95	0.95	0.586
Uneasiness	−2.19	2.39	0.361	−2.71	2.56	0.288	0.53	0.78	0.498
Depressive symptoms	**−3.39**	0.66	<0.001	**−4.31**	0.62	<0.001	0.91	0.59	0.125
Gastrointestinal issues	**−3.46**	0.92	<0.001	**−3.86**	0.93	<0.001	0.39	0.64	0.541

*b_{G9} = Gender differences before the reform; b_{G8} = Gender differences after the reform. Positive values indicate higher values for boys. Δ*b* = Difference of gender differences before (G9) minus after (G8) the reform. The metric of the latent variable was transformed to M = 500 and SD = 100 for standardized test performance and to M = 50 and SD = 10 for all other outcomes using pooled means and standard deviations. Statistically significant differences ($p < 0.05$) are printed in bold. Regression coefficients (b's) are based on group mean differences in a multiple group model. Two-sided p-values are reported. In cases where we had a directional hypothesis based on prior literature (e.g., higher stress scores of girls), one-sided p-values should be calculated/interpreted, which can be calculated by dividing the reported two-sided p-value by 2. Covariates that were considered for adjustment can be found in the Instrument section. For achievement, explained variance of the latent variables ranged between 12 and 44% (M = 26%), for self-concept between 1 and 15% (M = 6%), for interest between 1 and 11% (M = 4%), for stress between 0 and 6% (M = 2%), and for health between 0 and 13% (M = 5%). Note that when excluding course level as a covariate, results remained comparable regarding the size of estimates. In addition, in these models all SEs in the adjusted models were smaller compared to the unadjusted models. Please be aware that when estimating results for stress and health, we applied exploratory SEMs, which led to slightly different measurement models (i.e., differences in factor loadings), when additional variables (e.g., covariates) were considered and this explains differences in SEs between the adjusted and unadjusted solution. Although the general loading pattern (see **Supplementary Tables 3**, **4**) remained similar in adjusted and unadjusted models, SEs should not be directly compared across these two solutions, because they refer to slightly different latent variables.*

Taken as a whole, this study has brought to light several important findings. First of all, we found substantial gender disparities in favor of boys at the end of upper secondary school on the respective STEM outcomes. Disparities were pronounced regarding the achievement in mathematics and physics and substantially smaller in biology (e.g., only 1/4 of the size of mathematics achievement). This is an important finding and underscores that gender-related disparities reported in prominent large-scale studies of students in Grade 9 might not reflect actual disparities at the end of upper secondary school in Germany, a key stage in the education system, right before students enroll in university. It also reflects previously articulated heterogeneity in disparities across countries (OECD, 2019; Parker et al., 2020) and underlines the importance of more closely considering disparities at different time points in the education system in future studies.

Second, our findings show that a unidimensional perspective on school-related stress and health masks result patterns that appeared when investigating the constructs at a more fine-grained level of underlying dimensions. A five-factor multiple group ESEM model constituted multidimensional school-related stress, and a six-factor model constituted health. Although the patterns were more or less consistent and in disadvantage of girls, there were exceptions, for instance regarding the Malaise aspect of school-related stress where we found disadvantages for boys, and on the Uneasiness aspect of health where we found no statistically significant differences.

Finally, and most important in the context of this study, the gender disparities evident before the reform seemed to perpetuate after the reform for STEM-related standardized test performance, self-concept, and interest. For school-related stress and health we found some statistically significant

gender × reform interaction effects more in line with an accumulated advantages/disadvantages model (see **Figure 1**; i.e., on the Overload dimension of stress, and the Overburdening and Achievement-related fear dimensions of health). This suggests that although both girls and boys reported substantially higher stress levels and lower health after the reform, the increase or decrease, respectively, was somewhat larger for girls than boys, at least on some stress and health facets.

Gender Disparities and the School Time Reform

As outlined above, we found large disparities between girls and boys at the end of upper secondary school on STEM-related outcomes. In most cases, these disparities followed stereotypical patterns: Overall, girls performed less well on standardized tests in math-intensive STEM subjects. In addition, girls reported lower self-concept and interest than boys in mathematics and physics, whereas there were no significant gender-related disparities in biology. When integrating our findings into the theoretical model (see **Figure 1**), we can summarize that in most cases we found evidence for the perpetuation model. Disparities before the reform on the respective outcomes were pronounced, and these differences did not change much after the reform. Our findings extend prior findings in three regards: They are based on a later period in the education system (end of upper secondary school, right before the transition to university), a broadened set of outcomes, and a more fine-grained investigation of school-related stress and health.

As we outlined in the theoretical background, several prior studies had suggested treatment effect heterogeneity for high and low achievers (e.g., Nomi and Allensworth, 2009; Lavy, 2015; Huebener et al., 2017), which is why we expected we would find a pattern of results in line with the accumulated advantages model (**Figure 1**) for STEM outcomes. However, aside from few stress and health facets, we did not find any changes when comparing gender disparities before and after the reform. This might have had different causes—for instance, students in our sample were older at the end of secondary school, compared with students in the reviewed studies. Therefore, our sample might constitute a positive selection of higher performing students as some lower performing students might have dropped out before or in early upper secondary school or might have switched to vocational upper secondary schools, where this reform was not implemented. This might have led to smaller gender differences in upper secondary school than before, in lower secondary school. Further, the major changes of the G8-reform happened in lower secondary school, whereas upper secondary school remained largely unaffected. Therefore, potential interaction effects on STEM outcomes might already have "washed out" by the end of upper secondary school. Most importantly, when comparing differences between G8 and G9 students' average weekly hours spent in STEM courses, we found negligible differences. This means, that changes in subject-specific instructional time might have been a too small and a central factor for why we did not find any differences on STEM related outcomes. However, this would not explain previously found reform-specific differences between G8 and G9 students for instance in Biology (Hübner et al., 2017a).

In contrast to perpetuating disparities on STEM outcomes after the reform, our study revealed some statistically significant interaction effects on school-related stress and health. Importantly, both girls and boys tended to report more school-related stress and health issues after the reform. However, we did not find interaction effects on all stress and health dimensions, but only on those more related to school, namely the Overload dimension of school-related stress, and the Overburdening and Achievement-related fear dimensions of health. Compared to the perpetuating subject-specific results outlined above, these findings are slightly more in line with the proposed accumulated (dis)advantages model: On average, all students (girls and boys) tended to report higher stress/poorer health after the reform, but particularly those students who were more stressed/had lower health scores before the reform seemed to experience higher school-related stress and poorer health afterward, at least on stress and health facets more closely related to school. These results are in line with prior findings that girls report lower wellbeing scores than boys (e.g., Moksnes et al., 2010; Salmela-Aro and Tynkkynen, 2012; Tuominen-Soini and Salmela-Aro, 2014) and reflect findings from prior studies that students might perceive the remaining leisure time to be too limited to recover from school-related stress (Milde-Busch et al., 2010). The higher average workload per week in lower secondary school as a result of the G8-reform might have been one driver of the unevenly higher stress for girls after the reform. Other potentially relevant stressors than the higher workload could have included longer school days, the abolishment of Grade 11, or completing the same curriculum in a shorter amount of time. However, we cannot trace back which stressors might have ultimately fostered these results, as all of these potential causes are perfectly confounded with the reform (i.e., all changes happened simultaneously), we cannot disentangle their effects.

Limitations

There are several limitations that are important to consider when interpreting the results of this study. These limitations include potential threats to internal and external validity. Regarding internal validity, it is important to consider that we used data from a cohort control design, whereby two representative cohorts of students from before the reform were compared with two representative cohorts of students after the reform. Although this cohort control design has been discussed as providing a good foundation for the investigation of intervention effects, as it resembles a natural experiment setting (Shadish et al., 2002), it might be possible that the cohorts already differed independent of the reform (e.g., due to historical events). In other words, we did not have a control group who did not receive the treatment at the same time that the students in the treatment group received the treatment (a difference in difference design; e.g., Cunningham, 2021). This of course provides a challenge for all research using reform data because reforms are typically implemented at the same time for all students in a specific state. Therefore, researchers are typically required to consider students from different states or cohorts within the same state

(before the reform) as control groups, which in turn introduces different challenges and assumptions, particularly regarding their comparability. To address this potential limitation, we used survey weights to assure representativeness of the different cohorts. Notably, response rates on all assessments were 90% or larger at the student level (e.g., IEA, 2013). In addition, we inspected potential differences between the cohorts and specified adjusted models, in which we controlled for important (presumably relatively time-stable) covariates. All those checks suggested that if selection bias was present in our study, it should have been small at most (e.g., Hübner et al., 2017a).

Furthermore, it is important to underline that our findings are based on self-reports and that we did not have more objective markers to assess stress and health, for instance using data from health insurance agencies, medication records, or cortisol measures. Therefore, it cannot be ruled out that students, at least in part, also reported feeling more stressed because of ongoing discussions with their parents, friends from G9 cohorts, or the media. However, even if part of this effect could have been explained by these aspects, the remaining differences would have still remained of practical significance (e.g., Milde-Busch et al., 2010; Hübner et al., 2017a; Quis, 2018).

Regarding external validity, it is important to keep in mind that we considered representative data of one specific reform in one specific German state (Baden-Württemberg). Therefore, the findings should be generalized cautiously to discussions about effects of changes in instructional time. Most importantly, as shown in prior studies (Else-Quest et al., 2010; OECD, 2019), results on gender-related disparities are very heterogenous in STEM subjects across countries. The authors argue that one of the main drivers of gender differences are differential opportunity structures (e.g., equity in school enrollment). Based on this, it remains to be shown if our findings can be generalized to other countries where gender disparities are less or even more strongly pronounced, compared to Germany. However, doing this would require similar reforms to be implemented in other countries, which we are not aware of, even after consulting a large reform database (OECD, 2015). This also becomes evident when inspecting further related literature. Among others, findings on this topic are based on quite heterogeneous reforms (e.g., Allensworth et al., 2009; Domina et al., 2015; Huebener et al., 2017; Marcus et al., 2020), based on randomized controlled trials (e.g., Meyer and van Klaveren, 2013; Andersen et al., 2016) or cross-sectional secondary data analysis (e.g., Lavy, 2015). Before generalizing results from our study to the general debate about learning time or other environments (e.g., other states or reforms), researchers and practitioners should carefully consider potential similarities and differences.

Finally, the major change implemented by the G8-reform constitutes a school time compression, which was implemented by increasing average time per week spent in lower secondary school (Homuth, 2017). However, beyond these changes, other different elements changed simultaneously with the introduction of the G8-reform, for instance, educational standards were introduced and schools were required to develop a school-specific curriculum (Hübner et al., 2017a). Therefore, although the instructional time change is probably the most dominant feature of the reform, we cannot rule out that other changes might have affected our findings. Results of our study should therefore be interpreted cautiously as reform effects (e.g., a combination of different changes happening at the same time) rather than as pure effects of a change in instructional time.

CONCLUSION

In this study, we investigated the gender-specific effects of an instructional school time reform on student achievement and motivation in STEM subjects, as well as on school-related stress and health. For most outcomes, we found substantial gender disparities favoring boys (e.g., in mathematics and physics), which did not intensify after the reform, but rather seemed to perpetuate. In contrast to subject-specific effects, significant gender × reform interaction effects were only evident on aspects of school-related stress and health, namely the Overload dimension of stress and the Overburdening and Achievement-related fear dimensions of health. From a more general standpoint our findings underscore the relevance of explicitly considering gender disparities when developing, implementing, and evaluating policy reforms.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: this manuscript uses data from the National Educational Panel Study (NEPS): Additional Study Baden-Wuerttemberg, doi: 10.5157/NEPS:BW:3.2.0. From 2008 to 2013, NEPS data were collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS has been carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

ETHICS STATEMENT

The NEPS study is conducted under the supervision of the German Federal Commissioner for Data Protection and Freedom of Information (BfDI) and in coordination with the German Standing Conference of the Ministers of Education and Cultural Affairs (KMK) and – in the case of surveys at schools – the Educational Ministries of the respective Federal States. The studies involving human participants, including all data collection procedures, instruments, and documents, were reviewed and approved by the data protection unit of the Leibniz Institute for Educational Trajectories (LIfBi). Written informed consent to participate in this study was provided by the participants, if they were 18 years or older or the participants, and their legal guardian/next of kin, if they were below 18 years old (18 is the legal age of consent in Germany). The necessary steps are taken to protect participants' confidentiality according to national and international regulations of data security. Participation in the NEPS study is voluntary and

based on the informed consent of participants. This consent to participate in the NEPS study can be revoked at any time.

AUTHOR CONTRIBUTIONS

NH: conceptualization, formal analysis, writing—original draft, writing—review and editing, and project administration. WW, JM, and HW: conceptualization, writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

We acknowledge support by the Open Access Publishing Fund of University of Tübingen.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.816358/full#supplementary-material

REFERENCES

Adams, R. J. (2005). Reliability as a measurement design effect. *Stud. Educ. Eval.* 31, 162–172. doi: 10.1016/j.stueduc.2005.05.008

Agnafors, S., Barmark, M., and Sydsjö, G. (2021). Mental health and academic performance: a study on selection and causation effects from childhood to early adulthood. *Soc. Psychiatry Psychiatr. Epidemiol.* 56, 857–866. doi: 10.1007/s00127-020-01934-5

Allensworth, E., Nomi, T., Montgomery, N., and Lee, V. E. (2009). College preparatory curriculum for all: academic consequences of requiring algebra and English I for ninth graders in Chicago. *Educ. Eval. Policy Anal.* 31, 367–391. doi: 10.3102/0162373709343471

Altuzarra, A., Gálvez-Gálvez, C., and González-Flores, A. (2021). Is gender inequality a barrier to economic growth:A panel data analysis of developing countries. *Sustainability* 13:367. doi: 10.3390/su13010367

Andersen, S. C., Humlum, M. K., and Nandrup, A. B. (2016). Increasing instruction time in school does increase learning. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7481–7484. doi: 10.1073/pnas.1516686113

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., and vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: long-term reciprocal effects across five waves and three achievement tracks. *J. Educ. Psychol.* 109, 621–634. doi: 10.1037/edu0000163

Baumert, J., Nagy, G., and Lehmann, R. (2012). Cumulative advantages and the emergence of social and ethnic inequality: matthew effects in reading and mathematics development within elementary schools? *Child Dev.* 83, 1347–1367. doi: 10.1111/j.1467-8624.2012.01779.x

Bergmüller, S. (2003). "Schulische Belastung und gesundheitliche Beschwerden [School-related demands and health problems]," in *PISA 2003*, eds G. Haider and C. Reiter (Leykam: Internationaler Vergleich von Schülerleistungen).

Bergmüller, S. (2007). *Schulstress Unter Jugendlichen: Entstehungsbedingungen, vermittelnde Prozesse und Folgen. Eine empirische Studie im Rahmen von PISA 2003 [School stress among adolescents: Originating conditions, mediating processes, and consequences. An empirical study in the context of PISA 2003]*. Cham: Springer.

Berkowitz, T., Schaeffer, M. W., Maloney, E. A., Peterson, L., Gregor, C., Levine, S. C., et al. (2015). Math at home adds up to achievement in school. *Science* 350, 196–198. doi: 10.1126/science.aac7427

Biewen, M., and Schwerter, J. (2021). Does more maths and natural sciences in high school increase the share of female STEM workers? Evidence from a curriculum reform. *Appl. Econ.*. doi: 10.1080/00036846.2021.1983139 [Epub ahead of print]

Blossfeld, H. P., Rossbach, H. G., and von Maurice, J. (eds) (2011). *Education as a Lifelong Process - The German National Educational Panel Study (NEPS)*. Wiesbaden: Springer VS.

Büttner, B., and Thomsen, S. L. (2015). Are we spending too many years in school? Causal evidence of the impact of shortening secondary school duration. *German Econ. Rev.* 16, 65–86. doi: 10.1111/geer.12038

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Cuban, L. (2008). The perennial reform: fixing school time. *Phi Delta Kappan* 90, 240–250. doi: 10.1177/003172170809000404

Cunningham, S. (2021). *Causal Inference: The Mixtape*. London: Yale University Press.

Denissen, J. J. A., Zarrett, N. R., and Eccles, J. S. (2007). I like to do it, I'm able, and I know I am: longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Dev.* 78, 430–447. doi: 10.1111/j.1467-8624.2007.01007.x

DiStefano, C., and Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Struct. Equ. Model. Multidiscip. J.* 13, 440–464. doi: 10.1207/s15328007sem1303_6

Domina, T., McEachin, A., Penner, A., and Penner, E. (2015). Aiming high and falling short. *Educ. Eval. Policy Anal.* 37, 275–295. doi: 10.3102/0162373714543685

Duchhardt, C. (2015). *NEPS Technical Report for Mathematics: Scaling Results for the Additional Study Baden-Württemberg*. Bamberg: Leibniz-Institut für Bildungsverläufe e.V.

Eccles, J. S. (1983). "Expectancies, values, and academic behaviors," in *Achievement and Achievement Motivation*, ed. J. Spence (Dallas, TX: Freeman), 75–146.

Eccles, J. S., and Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annu. Rev. Psychol.* 53, 109–132. doi: 10.1146/annurev.psych.53.100901.135153

Eccles, J. S., and Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. *Contemp. Educ. Psychol.* 61:101859. doi: 10.1016/j.cedpsych.2020.101859

Else-Quest, N. M., Hyde, J. S., and Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychol. Bull.* 136, 103–127. doi: 10.1037/a0018053

Enders, C. K. (2010). *Applied Missing Data Analysis*. New York, NY: Guilford Press.

Fiorilli, C., de Stasio, S., Di Chiacchio, C., Pepe, A., and Salmela-Aro, K. (2017). School burnout, depressive symptoms and engagement: their combined effect on student achievement. *Int. J. Educ. Res.* 84, 1–12. doi: 10.1016/j.ijer.2017.04.001

Frenzel, A. C., Goetz, T., Pekrun, R., and Watt, H. M. G. (2010). Development of mathematics interest in adolescence: influences of gender, family, and school context. *J. Res. Adolesc.* 20, 507–537. doi: 10.1111/j.1532-7795.2010.00645.x

Haberkorn, K., and Pohl, S. (2013). *Cognitive basic skills – Data in the Scientific Use File*. Bamberg: University of Bamberg.

Hammond, A., Matulevich, E. R., Beegle, K., and Kumaraswamy, S. K. (2020). *The Equality Equation: Advancing the Participation of Women and Girls in STEM*. Washington, DC: The World Bank.

Hampel, P., and Petermann, F. (2006). Perceived stress, coping, and adjustment in adolescents. *J. Adolesc. Health* 38, 409–415. doi: 10.1016/j.jadohealth.2005.02.014

Hjern, A., Alfven, G., and Östberg, V. (2008). School stressors, psychological complaints and psychosomatic pain. *Acta Paediatr.* 97, 112–117. doi: 10.1111/j.1651-2227.2007.00585.x

Homuth, C. (2017). *Die G8-Reform in Deutschland [The G8-reform in Germany]*. Berlin: Springer.

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* 6, 1–55. doi: 10.1080/10705519909540118

Hübner, N., Rieger, S., and Wagner, W. (2016a). *NEPS Technical Report for Biological Competence: Scaling Results for the Additional Study*

Baden-Wuerttemberg (NEPS Survey Paper No. 9). Bamberg: Leibniz Institute for Educational Trajectories.

Hübner, N., Rieger, S., and Wagner, W. (2016b). *NEPS Technical Report for Physics Competence: Scaling Results for the Additional Study Baden-Wuerttemberg (NEPS Survey Paper No. 11)*. Bamberg: Leibniz Institute for Educational Trajectories.

Hübner, N., Wagner, W., Hochweber, J., Neumann, M., and Nagengast, B. (2020). Comparing apples and oranges: curricular intensification reforms can change the meaning of students' grades! Curricular intensification reforms can change the meaning of students' grades! *J. Educ. Psychol.* 112, 204–220. doi: 10.1037/edu0000351

Hübner, N., Wagner, W., Kramer, J., Nagengast, B., and Trautwein, U. (2017a). Die G8-Reform in Baden-Württemberg: kompetenzen, Wohlbefinden und Freizeitverhalten vor und nach der Reform: kompetenzen, Wohlbefinden und Freizeitverhalten vor und nach der Reform [The G8 reform in Baden-Württemberg: competencies, wellbeing and leisure time before and after the reform]. *Z. Erziehungswissenschaft* 20, 748–771. doi: 10.1007/s11618-017-0737-3

Hübner, N., Wille, E., Cambria, J., Oschatz, K., Nagengast, B., and Trautwein, U. (2017b). Maximizing gender equality by minimizing course choice options? Effects of obligatory coursework in math on gender differences in STEM. *J. Educ. Psychol.* 109, 993–1009. doi: 10.1037/edu0000183

Hübner, N., Wagner, W., Nagengast, B., and Trautwein, U. (2019). Putting all students in one basket does not produce equality: gender-specific effects of curricular intensification in upper secondary school. *Sch. Effect. Sch. Improv.* 30, 261–285. doi: 10.1080/09243453.2018.1504801

Huebener, M., Kuger, S., and Marcus, J. (2017). Increased instruction hours and the widening gap in student performance. *Labour Econ.* 47, 15–34. doi: 10.1016/j.labeco.2017.04.007

IEA (2013). *Methodenbericht: NEPS Zusatzstudie zur G8-Reform in Baden-Württemberg: Haupterhebung - Frühjahr 2011 (A72) [Technical report: NEPS additional study on the G8 reform in Baden-Württemberg*. Cham: Spring.

Jacobs, J. E., and Simpkins, S. D. (2005). Mapping leaks in the math, science, and technology pipeline. *New Dir. Child Adolesc. Dev.* 110, 3–6. doi: 10.1002/cd.145

Jöreskog, K. G., and Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *J. Am. Stat. Assoc.* 70, 631–639. doi: 10.1080/01621459.1975.10482485

Kühn, S. M., van Ackeren, I., Bellenberg, G., Reintjes, C., and Im Brahm, G. (2013). Wie viele Schuljahre bis zum Abitur? Eine multiperspektivische Standortbestimmung im Kontext der aktuellen Schulzeitdebatte [How many years until abitur in German upper secondary schooling? – Taking stock in the context of current school duration debates]. *Z. Erziehungswissenschaft* 16, 115–136. doi: 10.1007/s11618-013-0339-7

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *Econ. J.* 125, F397–F424. doi: 10.1111/ecoj.12233

Lazarides, R., Dicke, A. L., Rubach, C., and Eccles, J. S. (2020). Profiles of motivational beliefs in math: exploring their development, relations to student-perceived classroom characteristics, and impact on future career aspirations and choices. *J. Educ. Psychol.* 112, 70–92. doi: 10.1037/edu0000368

Lazarides, R., and Lauermann, F. (2019). Gendered paths into STEM-related and language-related careers: girls' and boys' motivational beliefs and career plans in math and language arts. *Front. Psychol.* 10:1243. doi: 10.3389/fpsyg.2019.01243

Ma, X., and Johnson, W. (2008). "Mathematics as the critical filter: curricular effects on gendered career choices," in *Gender and Occupational Outcomes: Longitudinal Assessments of Individual, Social, and Cultural Influences*, eds H. M. G. Watt and J. S. Eccles (Washington, DC: American Psychological Association), 55–83.

MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychol. Methods* 1, 130–149. doi: 10.1037/1082-989X.1.2.130

Maceira, H. (2017). Economic benefits of gender equality in the EU. *Intereconomics* 52, 178–183. doi: 10.1007/s10272-017-0669-4

Makarova, E., Aeschlimann, B., and Herzog, W. (2019). The gender gap in STEM fields: the impact of the gender stereotype of math and science on secondary students' career aspirations. *Front. Educ.* 4:60. doi: 10.3389/feduc.2019.00060

Marcus, J., Reif, S., Wuppermann, A., and Rouche, A. (2020). Increased instruction time and stress-related health problems among school children. *J. Health Econ.* 70:102256. doi: 10.1016/j.jhealeco.2019.102256

Marsh, H. W. (1990). The structure of academic self-concept: the Marsh/Shavelson model. *J. Educ. Psychol.* 82, 623–636. doi: 10.1037/0022-0663.82.4.623

Marsh, H. W. (1992). *Self Description Questionnaire (SDQ) III: A Theoretical and Empirical Basis for the Measurement of Multiple Dimensions of Late Adolescent Self-Concept: A Test Manual and a Research Monograph*. San Antonio, TX: The Psychological Corporation.

Marsh, H. W., and Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: beyond seductive pleasure and unidimensional perspectives. *Perspect. Psychol. Sci.* 1, 133–163. doi: 10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Martin, A. J., Yeung, A. S., and Craven, R. (2016). "Competence self-perceptions," in *Handbook of Competence and Motivation*, eds A. J. Elliot, C. S. Dweck, and D. Yeager (New York, NY: Guilford Press), 85–115.

Marsh, H. W., Morin, A. J. S., Parker, P. D., and Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annu. Rev. Clin. Psychol.* 10, 85–110. doi: 10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., and O'Neill, R. (1984). Self description questionnaire III: the construct validity of multidimensional self-concept ratings by late adolescents. *J. Educ. Meas.* 21, 153–174. doi: 10.1111/j.1745-3984.1984.tb00227.x

McNeish, D., Stapleton, L. M., and Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychol. Methods* 22, 114–140. doi: 10.1037/met0000078

Meinck, S., and Brese, F. (2020). *Gender Gaps in Science are Not a Given: Evidence on International Trends in Gender Gaps in Science Over 20 Years: Compass: Briefs in Education No. 11*. Paris: IEA.

Mejía-Rodríguez, A. M., Luyten, H., and Meelissen, M. R. M. (2021). Gender differences in mathematics self-concept across the world: an exploration of student and parent data of TIMSS 2015. *Int. J. Sci. Math. Educ.* 19, 1229–1250. doi: 10.1007/s10763-020-10100-x

Merton, R. K. (1968). The Matthew effect in science. *Science* 159, 56–63. doi: 10.1126/science.159.3810.56

Meyer, E., and van Klaveren, C. (2013). The effectiveness of extended day programs: evidence from a randomized field experiment in the Netherlands. *Econ. Educ. Rev.* 36, 1–11. doi: 10.1016/j.econedurev.2013.04.002

Meyer, T., Schneider, H., and Thomsen, S. L. (2019). New evidence on the effects of the shortened school duration in the German states: an evaluation of post-secondary education decisions. *German Econ. Rev.* 20, e201–e253. doi: 10.1111/geer.12162

Milde-Busch, A., Blaschek, A., Borggräfe, I., von Kries, R., Straube, A., and Heinen, F. (2010). Besteht ein Zusammenhang zwischen der verkürzten Gymnasialzeit und Kopfschmerzen und gesundheitlichen Belastungen bei Schülern im Jugendalter? [Is there an association between the reduced school years in grammar schools and headache and other health complaints in adolescent students?]. *Klinische Pädiatrie* 222, 255–260. doi: 10.1055/s-0030-1252012

Moksnes, U. K., Moljord, I. E., Espnes, G. A., and Byrne, D. G. (2010). The association between stress and emotional states in adolescents: the role of gender and self-esteem. *Pers. Individ. Diff.* 49, 430–435. doi: 10.1016/j.paid.2010.04.012

Muthén, L. K., and Muthén, B. O. (1998-2017). *Mplus User's Guide*, Eighth Edn. Los Angeles, CA: Muthén & Muthén.

Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., and Baumert, J. (2010). The development of students' mathematics self-concept in relation to gender: different countries, different trajectories? *J. Res. Adolesc.* 20, 482–506. doi: 10.1111/j.1532-7795.2010.00644.x

Nomi, T., and Allensworth, E. (2009). "Double-Dose" algebra as an alternative strategy to remediation: effects on students' academic outcomes. *J. Res. Educ. Effect.* 2, 111–148. doi: 10.1080/19345740802676739

OECD (2015). *Education Policy Outlook 2015*. Paris: OECD.

OECD (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. Paris: OECD Publishing.

Oehlert, G. W. (1992). A note on the delta method. *Am. Stat.* 46, 27–29. doi: 10.1080/00031305.1992.10475842

Parker, P. D., Schoon, I., Tsai, Y. M., Nagy, G., Trautwein, U., and Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: a multicontext study. *Dev. Psychol.* 48, 1629–1642. doi: 10.1037/a0029167

Parker, P. D., van Zanden, B., Marsh, H. W., Owen, K., Duineveld, J. J., and Noetel, M. (2020). The intersection of gender, social class, and cultural context: a meta-analysis. *Educ. Psychol. Rev.* 32, 197–228. doi: 10.1007/s10648-019-09493-1

Patall, E. A., Cooper, H., and Allen, A. B. (2010). Extending the school day or school year. *Rev. Educ. Res.* 80, 401–436. doi: 10.3102/0034654310377086

Pischke, J. S. (2007). The impact of length of the school year on student performance and earnings: evidence from the German short school years. *Econ. J.* 117, 1216–1242. doi: 10.1111/j.1468-0297.2007.02080.x

Quis, J. S. (2018). Does compressing high school duration affect students' stress and mental health? Evidence from the national educational panel study. *J. Econ. Stat.* 238, 441–476. doi: 10.1515/jbnst-2018-0004

Rawson, K., Stahovich, T. F., and Mayer, R. E. (2017). Homework and achievement: using smartpen technology to find the connection. *J. Educ. Psychol.* 109, 208–219. doi: 10.1037/edu0000130

Salmela-Aro, K., and Tynkkynen, L. (2012). Gendered pathways in school burnout among adolescents. *J. Adolesc.* 35, 929–939. doi: 10.1016/j.adolescence.2012.01.001

Santos Silva, M., and Klasen, S. (2021). Gender inequality as a barrier to economic growth: a review of the theoretical literature. *Rev. Econ. Household* 19, 581–614. doi: 10.1007/s11150-020-09535-6

Scheerens, J. (ed.) (2014). *Effectiveness of Time Investments in Education*. Berlin: Springer International Publishing.

Schönberger, B., and Aßmann, C. (2014). *Weighting the Additional Study in Baden-Wuerttemberg of the National Educational Panel Study*. Bamberg: University of Bamberg.

Schoon, I., and Eccles, J. S. (eds) (2014). *Gender Differences in Aspirations and Attainment: A Life Course Perspective*. Cambridge, MA: Cambridge University Press.

Seaton, M., Marsh, H. W., Parker, P. D., Craven, R. G., and Yeung, A. S. (2015). The reciprocal effects model revisited. *Gift. Child Q.* 59, 143–156. doi: 10.1177/0016986215583870

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Shavelson, R. J., Hubner, J. J., and Stanton, G. C. (1976). Self-concept: validation of construct interpretations. *Rev. Educ. Res.* 46, 407–441. doi: 10.3102/00346543046003407

Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S., and Haag, N. (2017). *IQB Trends in Student Achievement 2016: The Second National Assessment of German and Mathematics Proficiencies at the End of Fourth Grade*. Münster: Waxmann.

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., and Baumert, J. (2006). Tracking, grading, and student motivation: using group composition and status to predict self-concept and interest in ninth-grade mathematics. *J. Educ. Psychol.* 98, 788–806. doi: 10.1037/0022-0663.98.4.788

Trautwein, U., Neumann, M., Nagy, G., Lüdtke, O., and Maaz, K. (2010). "Institutionelle Reformen und individuelle Entwicklung: hintergrund und Fragestellung der Studie TOSCA-Repeat [Institutional reforms and individual development: background and research question of the TOSCA-Repeat study]," in *Schulleistungen von Abiturienten (1st ed.)*, eds U. Trautwein, M. Neumann, G. Nagy, O. Lüdtke, and K. Maaz (Berlin: Springer).

Tuominen-Soini, H., and Salmela-Aro, K. (2014). Schoolwork engagement and burnout among Finnish high school students and young adults: profiles, progressions, and educational outcomes. *Dev. Psychol.* 50, 649–662. doi: 10.1037/a0033898

United Nations General Assembly (1948). *Universal Declaration of Human Rights*. New York, NY: United Nations General Assembly.

Updegraff, K. A., Eccles, J. S., Barber, B. L., and O'brien, K. M. (1996). Course enrollment as self-regulatory behavior: who takes optional high school math courses? *Learn. Individ. Diff.* 8, 239–259. doi: 10.1016/S1041-6080(96)90016-3

van Sonderen, E., Sanderman, R., and Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: let's learn from cows in the rain. *PLoS One* 8:e68967. doi: 10.1371/journal.pone.0068967

Wagner, P., Schober, B., and Spiel, C. (2008). Time students spend working at home for school. *Learn. Instr.* 18, 309–320. doi: 10.1016/j.learninstruc.2007.03.002

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Dev.* 75, 1556–1574. doi: 10.1111/j.1467-8624.2004.00757.x

Watt, H. M. G., Bucich, M., and Dacosta, L. (2019). Adolescents' motivational profiles in mathematics and science: associations with achievement striving, career aspirations and psychological wellbeing. *Front. Psychol.* 10:990. doi: 10.3389/fpsyg.2019.00990

Watt, H. M. G., Hyde, J. S., Petersen, J., Morris, Z. A., Rozek, C. S., and Harackiewicz, J. M. (2017). Mathematics—a critical filter for STEM-related career choices? A longitudinal examination among Australian and U.S. adolescents. *Sex Roles* 77, 254–271. doi: 10.1007/s11199-016-0711-1

Watt, H. M. G., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., and Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: a comparison of samples from Australia, Canada, and the United States. *Dev. Psychol.* 48, 1594–1611. doi: 10.1037/a0027838

Widlund, A., Tuominen, H., and Korhonen, J. (2018). Academic well-being, mathematics performance, and educational aspirations in lower secondary education: changes within a school year. *Front. Psychol.* 9:297. doi: 10.3389/fpsyg.2018.00297

Wu, H., Guo, Y., Yang, Y., Le, Z., and Guo, C. (2021). A meta-analysis of the longitudinal relationship between academic self-concept and academic achievement. *Educ. Psychol. Rev.*. doi: 10.1007/s10648-021-09600-1 [Epub ahead of print]

Yu, C. Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes: A Dissertation Submitted in Partial Satisfaction of the Requirements for the Degree. Doctor of Philosophy in Education*. Los Angeles, CA: University of California.

Zhang, X., Noor, R., and Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PLoS One* 11:e0157795. doi: 10.1371/journal.pone.0157795

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 24 February 2022
doi: 10.3389/fpsyg.2022.846016

Gender Prejudice Within the Family: The Relation Between Parents' Sexism and Their Socialization Values

Daniela Barni[1]*, *Caterina Fiorilli*[2], *Luciano Romano*[2,3], *Ioana Zagrean*[2], *Sara Alfieri*[4] and *Claudia Russo*[2]

[1] Department of Human and Social Sciences, Università degli Studi di Bergamo, Bergamo, Italy, [2] Department of Human Sciences, Libera Università Maria SS. Assunta (LUMSA), Rome, Italy, [3] Department of Human Sciences, Universitá Europea di Roma (UER), Rome, Italy, [4] Department of Psychology, Università Cattolica del Sacro Cuore di Milano, Milan, Italy

Gender inequalities are still persistent despite the growing policy efforts to combat them. Sexism, which is an evaluative tendency leading to different treatment of people based on their sex and to denigration (hostile sexism) or enhancement (benevolent sexism) of certain dispositions as gendered attributes, plays a significant role in strengthening these social inequalities. As it happens with many other attitudes, sexism is mainly transmitted by influencing parental styles and socialization practices. This study focused on the association between parents' hostile and benevolent sexism toward women and their socialization values (specifically, conservation and self-transcendence), that are the values parents would like their children to endorse. We took both parents' and children's sex into account in the analyses. One-hundred-sixty-five Italian parental couples with young adult children participated in the study. Parents, both the mother and the father, individually filled in a self-report questionnaire composed of the Ambivalent Sexism Inventory and the Portrait Values Questionnaire. Findings showed that mothers' benevolent sexism was positively related to their desire to transmit conservation values to their sons and daughters. This result was also found for fathers, but with a moderation effect of children's sex. Indeed, the positive relationship between fathers' benevolent sexism and conservation was stronger in the case of sons than of daughters. Moreover, fathers' benevolent sexism was positively associated with self-transcendence values. Finally, fathers' hostile sexism was positively associated with conservation and negatively with self-transcendence. Limitations of the study, future research developments, and practical implications of the results are discussed.

Keywords: gender prejudice, hostile sexism, benevolent sexism, parents, socialization values

OPEN ACCESS

Edited by:
Elisabetta Sagone,
University of Catania, Italy

Reviewed by:
Orazio Licciardello,
University of Catania, Italy
Santa Parrello,
University of Naples Federico II, Italy

***Correspondence:**
Daniela Barni
daniela.barni@unibg.it

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *30 December 2021*
Accepted: *20 January 2022*
Published: *24 February 2022*

Citation:
Barni D, Fiorilli C, Romano L, Zagrean I, Alfieri S and Russo C (2022) Gender Prejudice Within the Family: The Relation Between Parents' Sexism and Their Socialization Values. Front. Psychol. 13:846016. doi: 10.3389/fpsyg.2022.846016

INTRODUCTION

Despite growing policy efforts designed to foster gender equality, culturally rooted and persistent inequalities are still around, and gender prejudice and sexism are thought to contribute significantly to this (Vandenbossche et al., 2018). Generally speaking, sexism is a form of prejudice and discrimination based on stereotypical beliefs about sex or gender (Dovidio et al., 2008). In

their Ambivalent Sexism Theory, Glick and Fiske (1996) innovatively considered sexism as a multidimensional construct composed of two sets of sexist attitudes, namely hostile and benevolent sexism. Both the forms of sexism fuel the subordination of women to men, although they deeply differ in their expression (Mastari et al., 2019). Hostile sexism refers to the traditional conceptualization of sexism as a reflection of hostility against women. Men are perceived as dominant over women, and the women who do not respect the conventional gender roles represent a potential threat to social order and men's power. Benevolent sexism, instead, is expressed in a seemingly positive and more subtle way. Women are paternalistically seen as loving but fragile individuals and therefore need men's protection and support. This protection is granted in exchange for women's respect of traditional gender roles (Glick and Fiske, 2001).

From childhood to young adulthood, parents play a key role in their children's development of gender-role attitudes and stereotypes (e.g., Halpern and Perry-Jenkins, 2016). Nevertheless, only a few empirical studies deal with hostile and benevolent sexism within the family. In general, the higher is parents' sexism, the stronger are their expectations that children behave in line with gender stereotypes. As a matter of fact, Garaigordobil and Aliri (2011) found a direct relationship between parents' hostile and benevolent sexist attitudes and their adolescent children's sexist attitudes, thus suggesting an intergenerational transmission of them (see also Montañés et al., 2012). The strength of this connection varied according to parents' and adolescents' sex, being higher between the mothers' and daughters' sexism and between the fathers' and sons' sexism. Effectively, parenting practices tend to reinforce gender-typed behaviors mainly, but not exclusively, within the same-sex parent-child dyads (e.g., Grusec and Goodnow, 1994; Lund et al., 2002). Lipowska et al. (2016), in their research concerning parental attitudes of couples with young children, showed the association between parents' sexism and parenting styles. The authors reported that fathers' sexism (both hostile and benevolent) was positively associated with inconsequence attitudes (i.e., unpredictable parenting behavior, mainly depending on parents' current mood) toward sons. In the case of daughters, fathers' hostile sexism supported overprotective attitudes, while the benevolent one was positively related to promoting autonomy. On the other side, mothers' benevolent sexism was negatively associated with overprotective and demanding attitudes toward sons, but not toward daughters. From Garaigordobil and Aliri (2012), which involved parental couples of adolescents, it turned out that parents' indulgent style (i.e., high involvement and low imposition) had the strongest relationship with a low level of adolescents' sexism (regardless of adolescents' sex).

Parents' socialization values are at the core of parenting styles and practices (e.g., Grusec and Goodnow, 1994; Kikas et al., 2014). Socialization values are the values parents would like their children to endorse (Barni et al., 2017), and they guide parents in raising and socializing their children both in the short-term (i.e., what values parents pursue for their children in the present) and in a long-term perspective (i.e., what values parents would like to see in their children in adulthood) (Lasker and Lasker, 1991; Tulviste et al., 2012). Previous studies mainly relied on Schwartz's Theory of basic human values (Schwartz, 1992, 2012) and showed that parents (both mothers and fathers) would like their sons and daughters to give importance to conservation values (i.e., tradition, conformity, and security) and self-transcendence values (i.e., benevolence and universalism) (Ranieri and Barni, 2012; Barni et al., 2017). Conservation and self-transcendence are both conceptualized as social-focused values because they mainly regulate the way people are socially related to others, relying on a principle of cooperation. However, they significantly differ from each other. On the one side, conservation values are self-protective values because they comply with the need to avoid conflicts, unpredictability, and changes. On the other side, self-transcendence values adhere to the need for relatedness, emphasizing the concern for the welfare of others and underlying self-expansive motivations (Schwartz, 2012; Russo et al., 2021).

The Present Study

Despite the relevance of both parents' sexism and socialization values in children's education and development, to the best of our knowledge, until now no studies have examined the association between them. This study aims to overcome this gap by analyzing the moderation effect of child's sex on the relation between parental sexism (i.e., hostile and benevolent sexism toward women) and the social-focused values (i.e., conservation and self-transcendence) parents would like to transmit to their young adult children.

The study involved Italian mothers and fathers. Italy is far from reaching satisfactory results in gender equality, despite relevant progress under the pressure of women's rights movements, civil society, and local and European legislation (Rosselli, 2014). In Italy, more and more young adults live with their parents for a long time. Young adulthood is an understudied stage of life concerning sexist socialization experiences, even though an increasing number of psychological studies have reported the important role of sexism in young romantic couples' birth, dynamics, and wellbeing (Lachance-Grzela et al., 2021).

We expected to find significant associations between parents' sexism and socialization values. In particular, we hypothesized that both hostile and benevolent sexism was positively associated with conservation values, which emphasize the importance of traditions and preservation of the status quo (Schwartz, 1992). On the contrary, we could hypothesize a negative relation between hostile sexism and self-transcendence values, emphasizing the importance of benevolence, gender equality, and social justice. It is, instead, not possible to make a sound hypothesis about the relation between benevolent sexism and self-transcendence. This is because, on the one side, benevolent sexism contributes to gender inequality and, on the other side, it promotes helping behaviors and intimate relationships. Moreover, given the absence of previous research on the topic, we did not formulate any specific hypotheses about the influence of parents' and children's sex on these associations.

METHOD

Participants and Procedure

One-hundred-sixty-five Italian married couples (mothers: M_{age} = 50.85, SD = 4.51; fathers: M_{age} = 53.98, SD = 5.47) with at least one young adult (M_{age} = 22.87, SD = 2.32) son (34.8%) or daughter (65.2%)[1] participated in the study, for a total of 330 participants. The couples were married for an average of 26.96 years (SD = 4.98) and lived in the North of Italy.

Parents were recruited through the collaboration of the universities attended by their young adult children. After being informed about the study nature and participants' rights, the parents who agreed to participate received two versions of an anonymous self-report questionnaire, one for the mother and one for the father. They completed them at home with the opportunity to phone researchers if any help was needed.

Measures

Sexism

The Ambivalent Sexism Inventory (Glick and Fiske, 1996, Italian adaptation by Manganelli Rattazzi et al., 2008) was used to measure parents' sexist attitudes. The scale is composed of 22 items on a 6-points Likert scale (0 = "Completely disagree"; 5 = "Completely agree") set into hostile sexism (item examples: "Women get offended too easily," "Most women fail to appreciate fully all that men do for them"; α_{mother} = 0.86; α_{father} = 0.88) and benevolent sexism (item examples: "Women have a superior moral sensibility," "Women should be cherished and protected by men"; α_{mother} = 0.81; α_{father} = 0.78).

Socialization Values

The subscales of conservation and self-transcendence values were extracted from the Portrait Values Questionaire (Schwartz et al., 2001) and adapted to measure parents' socialization values (Barni et al., 2017). Conservation includes 13 verbal portraits describing a person's goals, aspirations, or wishes that implicitly point to the importance of a value (item example: "She/he believes that people should do what they are told. She/he is convinced that people should always follow the rules, even when no one is checking"; α_{mother} = 0.85; α_{father} = 0.84). Self-transcendence includes 10 verbal portraits (item example: "It is very important for her/him to help the people around her/him. She/he aspires to take care of their wellbeing"; α_{mother} = 0.84; α_{father} = 0.86). Parents were asked to indicate their responses to the question: "How would you want your child to respond to each item?" on a 6-points Likert scale (1 = "Not like her/him" at all; 6 = "Very much like her/him").

Data Analysis

Preliminarily, we performed descriptive statistics of the study's variables and correlations between them. Then, we estimated four multiple hierarchical regression models to test the moderation effect of child's sex on the relations between mothers' and fathers' sexist attitudes and their socialization values. In the first two regression models, the outcome variables were mothers' conservation and self-transcendence, respectively. In the third and fourth models, the outcome variables were fathers' conservation and self-transcendence, respectively. In all the models, the independent variables were: children's age (Step 1), children's sex (1 = sons, 2 = daughters), parents' benevolent and hostile sexism (Step 2), and the interaction terms between parents' sexism and children's sex (Step 3). Before calculating the interaction terms, the single scores of continuous variables were centered on their means to reduce the risk of collinearity (Aiken and West, 1991).

The analyses were run using SPSS v.21.0 (George and Mallery, 2013) and Interaction! (Soper, 2010).

RESULTS

In **Table 1** descriptive statistics and correlations between the study's variables are reported.

Table 2 shows the results of the two hierarchical regression models referred to mothers' variables.

Only the association between mothers' benevolent sexism and conservation was statistically significant: the more mothers endorsed benevolent sexism, the more they wanted their sons and daughters to give importance to values such as tradition, conformity, and security. Children's sex did not moderate any associations between mothers' sexism and socialization values.

Table 3 contains the results of the two hierarchical regression models referred to fathers' variables.

Findings showed that fathers' hostile and benevolent sexist attitudes were positively related to conservation. Besides, fathers' hostile and benevolent attitudes were significantly related to self-transcendence, but in opposite directions (negative for hostile sexism and positive for benevolent sexism). Interestingly, children's sex moderated the relation between fathers' benevolent sexism and conservation values. As illustrated in **Figure 1**, the simple slope analysis revealed that the positive link between benevolent sexism and conservation was stronger in the case of sons [Simple slope = 0.37, SE = 0.12; 95% CI (0.13, 0.61), $p < 0.01$], than in the case of daughters [Simple slope = 0.12, SE = 0.07; 95% CI (−0.02, 0.27), $p > 0.05$].

DISCUSSION

The current study explored the association between parents' hostile and benevolent sexism toward women and socialization values. In particular, we considered the social-focused values (i.e., conservation and self-transcendence), which contribute to regulating how people relate socially to each other (Schwartz, 2012). We involved both mothers and fathers and analyzed the moderation effect of children's sex on the sexism-socialization values link.

The straightforward result is that parents' sexism is significantly associated with the social-focused values parents would like to see in their children. There are two related points to this main result: first, parents' benevolent sexism, more than the hostile one, seems to be involved in children's value socialization;

[1]If parents had more than one young-adult child, they were asked to respond thinking about their firstborn.

TABLE 1 | Descriptive statistics and correlations.

	M	SD	Min	Max	SK	K	3	4	5	6	7	8	9	10
1. CSEX	–	–	–	–	–	–	0.03	0.05	0.16*	−0.01	−0.03	0.05	−0.03	0.01
2. CAGE	22.87	2.32	20	31	0.95	0.86	−0.08	−0.17*	−0.03	−0.11	0.03	0.04	0.12	0.04
3. MHOSTSEX	2.01	0.92	0	4.10	−0.05	−0.54		0.64**	0.33**	0.17*	0.28**	0.09	−0.15	−0.18*
4. MBENSEX	2.32	0.90	0	4.60	−0.35	−0.14			0.27**	0.20*	0.34**	0.18*	−0.17*	−0.12
5. FHOSTSEX	2.50	0.95	0.10	5.00	−0.12	−0.31				0.27**	0.16*	0.23**	−0.14	−0.25**
6. FBENSEX	2.84	0.82	0.70	4.60	−0.38	−0.20					0.07	0.23**	0.03	0.10
7. MCONS	4.00	0.79	1.92	6.00	−0.27	−0.08						0.19*	0.38**	−0.08
8. FCONS	4.16	0.71	2.15	5.85	−0.10	−0.16							0.04	0.48**
9. MSELFT	4.90	0.60	3.10	6.00	−0.48	0.05								0.18*
10. FSELFT	4.81	0.68	2.80	6.00	−0.43	0.04								

*$p < 0.05$, **$p < 0.01$ (2-tails); M, Mean; SD, Standard deviation; SK, Skewness; K, Kurtosis; CSEX, Children's sex; CAGE, Children's age; MHOSTSEX, Mothers' hostile sexism; MBENSEX, Mothers' benevolent sexism; FHOSTSEX, Fathers' hostile sexism; FBENSEX, Fathers' benevolent sexism; MCONS, Mothers' conservation; FCONS, Fathers' conservation; MSELFT, Mothers' self-transcendence; FSELFT, Fathers' self-transcendence.*

TABLE 2 | Hierarchical multiple regression models with mothers' variables.

	Conservation values			Self-transcendence values		
	β	*t*	Model summary	*β*	*t*	Model summary
Step 1			$R^2 = 0.00$			$R^2 = 0.02$
CAGE	0.03	0.36	$F_{(1,154)} = 0.13$	0.12	1.53	$F_{(1,154)} = 2.34$
Step 2			$R^2 = 0.12$**			$R^2 = 0.04$
CSEX	−0.04	−0.51	$F_{(4,151)} = 5.40$	0.00	0.00	$F_{(4,151)} = 1.59$
MHOSTSEX	0.07	0.65	$\Delta R^2 = 0.12$	−0.07	−0.66	$\Delta R^2 = 0.02$
MBENSEX	0.31	3.04**		−0.11	−1.00	
Step 3			$R^2 = 0.13$**			$R^2 = 0.04$
MHOSTSEX *CSEX	0.45	1.10	$F_{(6,149)} = 3.84$	−0.18	−0.41	$F_{(6,149)} = 1.12$
MBENSEX *CSEX	−0.50	−1.15	$\Delta R^2 = 0.01$	0.29	0.63	$\Delta R^2 = 0.00$

*$p < 0.05$, **$p < 0.01$, R^2 = R-square; ΔR^2 = R-square changes; CAGE, Children's age; CSEX, Children's sex; MHOSTSEX, Mothers' hostile sexism; MBENSEX, Mothers' benevolent sexism.*

TABLE 3 | Hierarchical multiple regression models with fathers' variables.

	Conservation values			Self-transcendence values		
	β	*t*	Model summary	*β*	*t*	Model summary
Step 1			$R^2 = 0.00$			$R^2 = 0.00$
CAGE	0.04	0.52	$F_{(1,154)} = 0.27$	0.04	0.51	$F_{(1,154)} = 0.26$
Step 2			$R^2 = 0.09$**			$R^2 = 0.11$**
CSEX	0.03	0.34	$F_{(4,151)} = 3.59$	0.07	0.93	$F_{(4,151)} = 4.73$
FHOSTSEX	0.16	2.00*	$\Delta R^2 = 0.09$	−0.32	−3.98**	$\Delta R^2 = 0.11$
FBENSEX	0.20	2.47*		0.21	2.65**	
Step 3			$R^2 = 0.13$**			$R^2 = 0.13$**
FHOSTSEX *CSEX	0.45	1.55	$F_{(6,149)} = 3.47$	0.47	1.62	$F_{(6,149)} = 3.62$
FBENSEX *CSEX	−0.65	−2.18*	$\Delta R^2 = 0.04$	−0.17	−0.59	$\Delta R^2 = 0.02$

*$p < 0.05$, **$p < 0.01$, R^2 = R-square; ΔR^2 = R-square changes; CAGE, Children's age; CSEX, Children's sex; FHOSTSEX, Fathers' hostile sexism; FBENSEX, Fathers' benevolent sexism.*

second, fathers' sexism, more than the mothers' one, intervenes in children's value socialization.

As hypothesized, mothers' and fathers' benevolent sexism was positively related to conservation values. Parents characterized by high levels of benevolent sexism might want to pass on to their children conservation values because the content of these values is in line with the desire for a stable context in which women are protected from harm (Sortheix and Schwartz, 2017). It is worthwhile noting that the relationship between benevolent sexism and conservation to transmit to sons and daughters is the only one significant for mothers. Benevolent sexism is the most subtle type of sexism, generally endorsed by both genders, which includes valuing feminine-stereotypes (Mastari et al., 2019). Conservation values are self-protective values, that serve to cope with anxiety due to uncertainty in the world by avoinding conflict (conformity) and maintaining the current order (tradition and security). Thus, the relation between mothers' benevolent sexism and the socialization values of conservation may express the desire to ensure a long-term safety for women (across generations) by the fulfillment of traditional gender roles.

Interestingly, as shown by the moderated regression analysis, the positive relation between fathers' benevolent sexism and conservation was moderated by children's sex, being stronger in the case of sons than of daughters. Thus, from the fathers' view, it is the task of men (i.e., sons) to preserve stability and safety in order to protect and support women. Furthermore, fathers' benevolent sexism was related to growth and self-expansive values (i.e., self-transcendence). Fathers with high levels of benevolent sexism would likely interpret their sexist attitude as a form of respect and care toward women instead of an attitude hindering women's freedom. Their sexism might assume the shape of paternalism, thus strengthening their view of being a caring person (Glick and Fiske, 2001). Hence, they may wish to transmit to their children generative values (Erickson, 1963) whose content is related to the concern

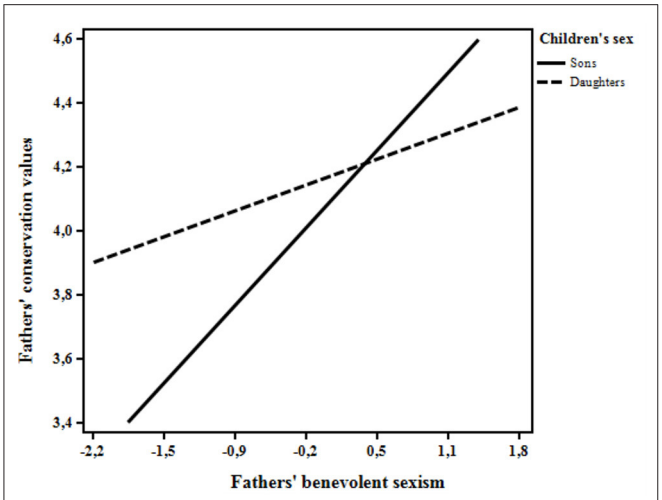


FIGURE 1 | The moderating role of children's sex on the relation between fathers' benevolent sexism and fathers' conservation values.

for the welfare and the protection of all human beings (Schwartz, 2012).

On the contrary, fathers' hostile sexism was not generative at all. It was positively associated with conservation values, but negatively with self-transcendence values. Men with high levels of hostile sexism tend to exhibit a hostile attitude toward women, reinforcing the view that women are only suited for domestic roles, even when women aspire to high-status roles that are perceived as suitable only for men (Eagly and Mladinic, 1994). As such, fathers' hostile sexism discourages the promotion of self-transcendence values that emphasize the understanding, appreciation, tolerance, and equality of all people.

Two main limitations of the present study must be acknowledged. First, the study's cross-sectional design did not allow us to draw causal interpretations from the results or catch potential changes over time. Second, the sample was of convenience and relatively small in size, with fewer sons than daughters. For these reasons, future longitudinal studies with larger representative sample of families are needed to better understand the role of parents' sexism within the family socialization processes.

Despite its limitations, this is the first study showing that parents' sexism intervenes at the core of socialization of young adult children by being related to what parents would like their children to value. There is a direct transmission of sexist attitudes between parents and children (Garaigordobil and Aliri, 2011) and an indirect path through promoting desired values across generations. Values represent, to some extent, a family heritage (e.g., Fiorilli et al., 2015). Parents have a mental representation of an "ideal adult," developed based on their own values, beliefs, and (sexist) attitudes that shape what they consider beneficial and adaptive. When they state the desired values for their children, they project such representation onto them (Rosenthal and Roer-Strier, 2006; Barni et al., 2017).

All in all, our results highlighted the "pervasive" role of fathers' sexism in children's value socialization. These results align with previous research showing the stronger influence of fathers' hostile and benevolent sexism on family relationships and dynamics (e.g., aggressive parenting, Overall et al., 2021). In the socialization of sexism, this seems especially true for the father-son dyad as suggested by our and previous studies (e.g., Garaigordobil and Aliri, 2011).

This study's findings can have significant practical implications. Interventions to reduce sexism are quite rare, practically absent in working with parents. Differently from other forms of prejudice (e.g., racial), intergroup contact cannot be applied to reduce sex prejudice. Providing individuals with gender-relevant information could be a good starting point to change their sexist attitudes (Becker and Swim, 2011). In this line, it would be helpful to develop training programs involving both mothers and fathers to strengthen parents' awareness about their hostile and benevolent sexist attitudes in order to avoid directly or indirectly transmitting them to future generations. We must not forget that cultural persistence is essentially the result of family and social transmission (Schönpflug and Bilz, 2009).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The study was reviewed and approved by the Scientific Board of the Family Studies and Research University Centre, Catholic University of Milan, and by the Ethical Committee of the Catholic University of Milan, Department of Psychology, Italy. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and international committees on human experimentation. All participants gave their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DB designed the study, collected the data, and contributed to the writing of the manuscript. SA designed the study, collected the data, and edited the manuscript. LR and IZ analyzed the data and wrote the Methods and Results. CF and CR contributed to the writing of Introduction and Discussion. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by research funds granted to the DB by the University of Bergamo.

REFERENCES

Aiken, L. S., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: SAGE Publications.

Barni, D., Ranieri, S., Donato, S., Tagliabue, S., and Scabini, E. (2017). Personal and family sources of parents' socialization values: A multilevel study. *Av. Psicol. Latinoam.* 35, 9–22. doi: 10.12804/revistas.urosario.edu.co/apl/a.3468

Becker, J. C., and Swim, J. K. (2011). Seeing the unseen: Attention to daily encounters with sexism as way to reduce sexist beliefs. *Psychol. Women Q.* 35, 227–242. doi: 10.1177/0361684310397509

Dovidio, J. F., Glick, P., and Rudman, L. A. (2008). *On the Nature of Prejudice: Fifty Years After Allport.* Hoboken, NJ: Blackwell Publishing Ltd.

Eagly, A. H., and Mladinic, A. (1994). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *Eur. Rev. Soc. Psychol.* 5, 1–35. doi: 10.1080/14792779543000002

Erickson, E. H. (1963). *Childhood and Society*. New York, NY: WW Norton & Company.

Fiorilli, C., De Stasio, S., Di Chicchio, C., and Chan, S. M. (2015). Emotion socialization practices in Italian and Hong Kong-Chinese mothers. *Springerplus* 4, 1–9. doi: 10.1186/s40064-015-1550-1

Garaigordobil, M., and Aliri, J. (2011). Conexión intergeneracional del sexismo: influencia de variables familiares. *Psicothema* 23, 382–387. doi: 10.6018/analesps.33.2.232371

Garaigordobil, M., and Aliri, J. (2012). Parental socialization styles, parents' educational level, and sexist attitudes in adolescence. *Span. J. Psychol.* 15, 592–603. doi: 10.5209/rev_SJOP.2012.v15.n2.38870

George, D., and Mallery, P. (2013). *IBM SPSS Statistics 21 Step by Step: A Simple Guide and Reference*. New York, NY: Pearson.

Glick, P., and Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *J. Pers. Soc. Psychol.* 70, 491–512. doi: 10.1037/0022-3514.70.3.491

Glick, P., and Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *Am. Psychol.* 56, 109–118. doi: 10.1037/0003-066X.56.2.109

Grusec, J. E., and Goodnow, J. J. (1994). Impact of parental discipline methods on the child's internalization of values: A reconceptualization of current points of view. *Dev. Psychol.* 30, 4–19. doi: 10.1037/0012-1649.30.1.4

Halpern, H. P., and Perry-Jenkins, M. (2016). Parents' gender ideology and gendered behavior as predictors of children's gender-role attitudes: A longitudinal exploration. *Sex Roles* 74, 527–542. doi: 10.1007/s11199-015-0539-0

Kikas, E., Tulviste, T., and Peets, K. (2014). Socialization values and parenting practices as predictors of parental involvement in their children's educational process. *Early Educ. Dev.* 25, 1–18. doi: 10.1080/10409289.2013.780503

Lachance-Grzela, M., Liu, B., Charbonneau, A., and Bouchard, G. (2021). Ambivalent sexism and relationship adjustment among young adult couples: An actor-partner interdependence model. *J. Soc. Pers. Relat.* 38, 2121–2140. doi: 10.1177/02654075211005549

Lasker, A. A., and Lasker, J. N. (1991). Do they want their children to be like them?: Parental heritage and Jewish identity. *Contemp. Jew.* 12, 109–126. doi: 10.1007/BF02965537

Lipowska, M., Lipowski, M., and Pawlicka, P. (2016). Daughter and son: A completely different story? Gender as a moderator of the relationship between sexism and parental attitudes. *Heal. Psychol. Rep.* 4, 224–236. doi: 10.5114/hpr.2016.62221

Lund, L. K., Zimmerman, T. S., and Haddock, S. A. (2002). The theory, structure, and techniques for the inclusion of children in family therapy: A literature review. *J. Marital Fam. Ther.* 28, 445–454. doi: 10.1111/j.1752-0606.2002.tb00369.x

Manganelli Rattazzi, A. M., Volpato, C., and Canova, L. (2008). L'atteggiamento ambivalente verso donne e uomini. Un contributo alla validazione delle scale ASI e AMI [Ambivalent attitudes toward women and men. A contribution to the validation of the ASI and AMI scales]. *G. Ital. Psicol.* 35, 217–246. doi: 10.1421/26601

Mastari, L., Spruyt, B., and Siongers, J. (2019). Benevolent and hostile sexism in social spheres: The impact of parents, school and romance on Belgian adolescents' sexist attitudes. *Front. Sociol.* 4, 47. doi: 10.3389/fsoc.2019.00047

Montañés, P., de Lemus, S., Bohner, G., Megías, J. L., Moya, M., and Garcia-Retamero, R. (2012). Intergenerational transmission of benevolent sexism from mothers to daughters and its relation to daughters' academic performance and goals. *Sex Roles* 66, 468–478. doi: 10.1007/s11199-011-0116-0

Overall, N. C., Chang, V. T., Cross, E. J., Low, R. S. T., and Henderson, A. M. E. (2021). Sexist attitudes predict family-based aggression during a COVID-19 lockdown. *J. Fam. Psychol.* 35, 1043–1052. doi: 10.1037/fam0000834

Ranieri, S., and Barni, D. (2012). Family and other social contexts in the intergenerational transmission of values. *Fam. Sci.* 3, 1–3. doi: 10.1080/19424620.2012.714591

Rosenthal, M. K., and Roer-Strier, D. (2006). "What sort of an adult would you like your child to be?": Mothers' developmental goals in different cultural communities in Israel. *Int. J. Behav. Dev.* 30, 517–528. doi: 10.1177/0165025406072897

Rosselli, A. (2014). *The Policy on Gender Equality in Italy*. Available online at: https://www.europarl.europa.eu/RegData/etudes/note/join/2014/493052/IPOL-FEMM_NT(2014)493052_EN.pdf (accessed December 27, 2021).

Russo, C., Barni, D., Zagrean, I., Lulli, M. A., Vecchi, G., and Danioni, F. (2021). The resilient recovery from substance addiction: The role of self-transcendence values and hope. *Mediterr. J. Clin. Psychol.* 9, 1–20. doi: 10.6092/2282-1619/mjcp-2902

Schönpflug, U., and Bilz, L. (2009). "The transmission process: mechanisms and contexts," in *Cultural Transmission: Psychological, Developmental, Social, and Methodological Aspects. Culture and Psychology,* ed U. Schönpflug (Cambridge: Cambridge University Press), 212–239. doi: 10.1017/CBO9780511804670.011

Schwartz, S. H. (1992). "Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries," in *Advances in Experimental Social Psychology,* ed M. P. Zanna (Cambridge, MA: Academic Press), 1–65. doi: 10.1016/S0065-2601(08)60281-6

Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online Read. Psychol. Cult.* 2, 919–2307. doi: 10.9707/2307-0919.1116

Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., and Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *J. Cross. Cult. Psychol.* 32, 519–542. doi: 10.1177/0022022101032005001

Soper, D. (2010). *Interaction! Unpublished Copyrighted Software.* Available online at: https://www.danielsoper.com/Interaction/default.aspx. (accessed December 12, 2021).

Sortheix, F. M., and Schwartz, S. H. (2017). Values that underlie and undermine well–being: Variability across countries. *Eur. J. Pers.* 31, 187–201. doi: 10.1002/per.2096

Tulviste, T., Mizera, L., and De Geer, B. (2012). Socialization values in stable and changing societies: A comparative study of Estonian, Swedish, and Russian Estonian mothers. *J. Cross. Cult. Psychol.* 43, 480–497. doi: 10.1177/0022022111401393

Vandenbossche, L., Spruyt, B., and Keppens, G. (2018). Young, innocent and sexist? Social differences in benevolent and hostile sexist attitudes towards women amongst Flemish adolescents. *Young* 26, 51–69. doi: 10.1177/1103308817697240

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Education

ORIGINAL RESEARCH
published: 17 March 2022
doi: 10.3389/feduc.2021.780815

Do Gender Conformity Pressure and Occupational Knowledge Influence Stereotypical Occupation Preferences in Middle Childhood?

Stephanie Masters and Joan Barth*

University of Alabama, Tuscaloosa, AL, United States

This study investigates how perceived occupational knowledge, gender stereotypes, and pressure to conform to gender norms influence children's career interests in a sample of fourth and fifth grade children (n = 178, M_{age} = 9.78 years, 46.6% girls). Children were interested in and perceived that they knew more about own gender dominated occupations, compared to other gender dominated occupations. Gender moderated the effect of gender conformity pressure and gender stereotypes on interest in female-dominated but not male-dominated occupations. Boys were less interested in female-dominated occupations when they felt pressure to conform to gender norms and held more stereotypical beliefs about those occupations. These results suggest that perceived occupational knowledge is an important, yet overlooked, factor in understanding gender differences in children's occupational interests.

Keywords: children, gender differences, gender roles, stereotypes, occupational interest

OPEN ACCESS

Edited by:
Caterina Fiorilli,
Libera Università Maria SS. Assunta University, Italy

Reviewed by:
Claudia Russo,
Libera Università Maria SS. Assunta, Italy
Eleonora Farina,
University of Milano-Bicocca, Italy

****Correspondence:***
Stephanie Masters
smasters@crimson.ua.edu

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *21 September 2021*
Accepted: *17 December 2021*
Published: *17 March 2022*

Citation:
Masters S and Barth J (2022) Do Gender Conformity Pressure and Occupational Knowledge Influence Stereotypical Occupation Preferences in Middle Childhood? Front. Educ. 6:780815. doi: 10.3389/feduc.2021.780815

INTRODUCTION

Middle childhood (7–12 years old) is a unique time to explore how gender influences occupational interests. Elementary aged children hold less rigid gendered attitudes than younger children in some domains, yet their behaviors and interests tend to be gender-typed (Blakemore, 2003). Occupational gender stereotypes from the larger culture impact children's occupational interest, which remain relatively stable from early adolescence to middle adulthood (Low et al., 2005). Thus, gender differences in children's early occupational interests may have a lasting impact. This study examines several socialization factors associated with the emergence of these gender differences, including perceived pressure from parents, teachers, and peers to conform to gender role norms, perceived knowledge about occupations, and gender stereotypes, and further extends previous research that has focused primarily on middle class samples to children attending lower SES (socioeconomic status) schools.

Gender socialization and stereotypes are central to three major theoretical perspectives of the development of career interests. Gottfredson. (1981) circumscription and compromise theory proposes that social experiences shape children's perceptions of gender-typical behavior and roles, and as children grow, they increasingly rule out occupations that are atypical for their gender. Middle childhood is a critical time point as older elementary aged children start to consider the social desirability of occupations for their own gender. Additionally, social cognitive career theory (Lent et al., 1994) and the expectancy value theory (Eccles et al., 1983) point to gender socialization and related gender stereotypes to explain gender differences in academic and

occupational interests, proposing that gender socialization and stereotypes affect interests through academic or career self-concept and efficacy.

The current study extends this work and considers that knowledge of careers develops outside the influence of key socializing agents and is gendered in other ways. Gender as a social category is salient to children and affects their attention to people in their world, even when there is little pressure to conform to gender norms. According to social role theory (Eagly, 1987) and social cognitive theory of gender development (Bussey and Bandura, 1999), youth naturally attend to same-gender role models, supporting the idea that gendered knowledge of careers may emerge independent of social pressure from others. Gender differences in occupation interests may emerge due to many factors, some of which are captured by gender socialization measures and some that occur outside of the influence of salient social agents. In this study, we examine if felt pressure to conform to gender norms and perceived knowledge contribute to explaining gender differences in career interests in addition to more frequently studied constructs, such as gender stereotypes. Additionally, because socioeconomic status is related to children's occupational aspirations (Weinger, 2000), and lower SES groups are less represented in previous research in this area, this study focused on children attending lower SES schools.

Gender Socialization

Gender socialization, or the messages an individual receives about what behaviors and roles are culturally appropriate for one's gender, is a powerful influence on career choice (Eccles, 1987; Lent et al., 1994). Early socialization about what occupations are considered appropriate for one's gender may be especially impactful on later interest (e.g., Antecol and Cobb-Clark, 2013). In the current study, gender socialization was conceptualized as children's felt pressure to conform to gender norms. Although there is a host of empirical evidence investigating the impact that felt pressure to conform to gender norms has on children's educational (Vantieghem et al., 2014) and psychosocial outcomes (Corby et al., 2007; Masters et al., 2020), less work has examined how felt pressure to conform to gender norms impacts children's occupational interests. It should be noted that girls often report feeling less pressure to conform to gender norms than boys (Egan and Perry, 2001; Masters et al., 2020), suggesting that there may be differences between boys and girls with respect to the degree to which social pressure plays a role in their career interests.

In line with social cognitive theory (Bussey and Bandura, 1999), this study examines parents, teachers, and peers as sources of gender socialization (Rice et al., 2013). Adults (Sullivan et al., 2018) and peers (Pascoe, 2012; Heinze and Horn, 2014) respond more negatively to children who engage in gender-atypical activities than those who engage in gender typical activities. Children under pressure to adhere to gender roles may not explore a wide range of options when deciding what interests to pursue (Bem, 1981; Bussey and Bandura, 1999) and in an effort to avoid negative evaluations from parents, teachers, and peers, may adopt gender typical interests. As such, it was hypothesized that youth who feel a strong degree of gender conformity pressure would be more interested in own-gender dominated occupations than those who report less gender conformity pressure.

Gender Stereotypes

Despite societal changes in men's and women's roles, gender stereotypes have persisted across the decades (Haines et al., 2016). Gender stereotypes impact the development of career interests in a number of ways, including discouraging people from choosing careers considered incongruous with their gender (Eccles, 2011). In accordance with social role theory (Eagly, 1987), occupational interests are correlated with gender stereotyping and work-force gender-segregation, such that boys and girls are more interested in careers they believe are predominately held by their own gender (Hayes et al., 2018). Gender differences in children's occupational interests parallel the adult work world, suggesting that the gender composition of occupations has significant intergenerational effects. In a cyclical manner, as occupational demographics change historically, gender stereotypes should evolve (Koenig and Eagly, 2014). Given that boys are more beholden to strict gender stereotypes than girls (Pauletti et al., 2017), it is important to consider the role that gender stereotypes play in gender differentiated career interests.

Occupational Knowledge

Occupational knowledge is the understanding of information about careers, such as the physical and mental requirements, time, or status (Schmitt-Wilson and Welsh, 2012). Children's occupational preferences are linked to fields in which they feel knowledgeable (Rohlfing et al., 2012; Schmitt-Wilson and Welsh, 2012; Hartung, 2015). The *perception* of occupational knowledge or the amount of knowledge children think they have (Rohlfing et al., 2012), and their *actual* knowledge (Watson and McMahon, 2005) are two distinct factors. Boys tend to think they have more knowledge of masculine occupations than girls (and vice versa), even though boys and girls might be quite comparable in their actual knowledge (Miller and Hayward, 2006). There are gender differences in perceived occupational knowledge, but not in actual knowledge (Ferrari et al., 2015), suggesting that gender stereotypes influence self-perceived knowledge. We focused on perceived occupational knowledge as a factor for understanding gender differences, hypothesizing that children's perception of having more knowledge about same-than other-gender dominated occupations would be positively related to their interest in same-gender occupations. Both perceived knowledge and interests are impacted by gender stereotypes embedded in the larger culture.

Socioeconomic Status

This study investigates gendered career interests within the parameters of perceived occupational knowledge, occupational gender stereotypes, and gender conformity pressure among children in grades 4 and 5. An additional objective of this research is to expand the study of career interests to children in lower socioeconomic status (SES) schools who are less often represented in this area of research. Children from lower SES households receive less information about work (Doyle, 2011),

perceive more barriers towards career attainment (Weinger, 2000), and aspire to less prestigious jobs (Howard and Walsh, 2011) than students from middle to high SES households. Youth from lower SES households have been included in recent work regarding felt pressure to conform to gender norms (Cook et al., 2019; Shroeder and Liben, 2020) but have been less represented in work examining gendered occupational stereotypes and interests (see Patterson, 2012, for an exception). Accordingly, all participants in the study attended Title 1 schools.

Research Aims and Hypotheses

The first aim of this study is to examine the association between perceived occupational knowledge and occupational interest to determine whether our lower SES sample replicates the findings from past research showing that perceived occupational knowledge predicts occupational interest (Rohlgfing et al., 2012). The relationship between occupational knowledge and interest has primarily been investigated in European samples (Rohlgfing et al., 2012). Our work expands upon current literature by examining this effect in a sample of lower SES students residing in the U.S.

The second aim is to examine gender differences in perceived occupational knowledge, occupational interest, and felt pressure to conform to gender norms. It was hypothesized that girls would have greater perceived occupational knowledge and interest in female-dominated occupations, compared to boys, and there would be comparable findings for boys for male-dominated occupations. In line with prior work, we expected that boys would experience gender conformity pressure to a greater degree than girls (Egan and Perry, 2001; Masters et al., 2020).

Most importantly the third aim is to examine the combined effects of perceived occupational knowledge, felt pressure to conform, and gender stereotypes on occupation interests. Analyses tested the hypothesis that perceived knowledge may capture unique variance relative to the other measures. These analyses also tested for gender differences in the effects of the three factors.

METHODS

Participants

Participants were 178 fourth and fifth grade students (53.4% boys, 53.9% fourth graders; M_{age} = 9.78 years, SD = 0.717; range 9–11) recruited from four local schools and after-school programs in the U.S. Southeast. All elementary schools were Title 1 schools, with a range of 55–69% of students qualifying for free/reduced lunch. The ethnic makeup of the final sample was 65.2% White, 23% Black, 2.2% Latinx/Hispanic, 3.4% Native American, and 5.6% were another race. This is similar to the racial demographics of the participating schools.

The initial response rate was 61.2% of the 374 potential students. Of the 229 parent consent forms returned, 211 (92.3%) gave consent. From this group 178 (77.7%) completed surveys. Four children declined to participate, and the rest were absent the day the survey was administered. An a priori power analysis (G*Power; Faul et al., 2007) was conducted with $\alpha = 0.05$ and power set to 0.80. To detect a small to medium effect size (0.15) a sample of 130 participants was needed, and the actual sample size was larger.

Procedure

Data collection occurred during the fall semester of school. Surveys were administered to all assenting students during school hours. Participants were informed that participation is optional, their responses would be kept confidential, and participation could be terminated at any time during the study, for any reason and without penalty. After reviewing the assent statement, students completed the paper survey on their own. Research staff were available to address questions and ensure that students stayed on task and did not share their answers. Teachers remained in the room during the survey as required by school district policy. However, they were not involved in the administration of the survey and were generally sitting at their desks away from students.

Measures

Knowledge, interest, and stereotype questions were asked for four male-dominated (Construction worker, Fire fighter, Engineer, Computer Programmer) and four female-dominated (Elementary School Teacher, Nurse, Librarian, Hair Stylist) occupations. Male-dominated and female-dominated occupations were selected based on current data from the Bureau of Labor Statistics (2016). Similar to Fulcher (2011), there was very little difference between the combined four-male and combined four-female occupations on salary and required education. Participants had to have answered 75% of the items on a scale to receive a score for that scale.

Occupational interest. Children were asked: "How much would you like to be a(n) (occupation)?" Participants responded using a 6-point scale (1 = not at all to 6 = very much). Summary scores of participants' interests in masculine and feminine occupations were computed by averaging responses to the four items of each type, with higher scores indicating a greater preference for feminine or masculine occupations. This measure was adapted from the Occupations, Activities and Traits - Personal Measure (Liben and Bigler, 2002). In the current study, reliability estimates were moderate for the four feminine items (ω = 0.66; α = 0.66) and the four masculine items (ω = 0.62; α = 0.61). The internal reliability for occupational interest is similar to those in other studies with children who were enrolled in fourth, fifth, or sixth grades (Spence and Hall, 1996; Barth et al., 2018; Pacilli et al., 2019).

Occupational knowledge. The Rohlfing et al. (2012) Occupational Knowledge Scale was adapted to measure perceived occupational knowledge. For each occupation, children rated "About how much do you already know about what people in this job do?" on a 6-point scale (1 = not very much; 6 = a lot). Scores were averaged to create a male-dominated (ω = 0.65; α = 0.66) and a female-dominated scale (ω = 0.69; α = 0.68). Each respective scale was four items each.

Occupation Gender Stereotyping. Similar to Liben and Bigler (2002), participants were asked "*who would like to have this job*" to evaluate to the extent to which each job is perceived as being gender segregated. Response options ranged

TABLE 1 | Bivariate correlations among measures for boys and girls.

Measure	1	2	3	4	5	6	7
Male-dominated occupations							
Knowledge	—	0.63***	0.05	0.31**	0.12	0.05	0.12
Interest	0.41***	—	−0.08	0.14	0.09	0.03	0.06
Stereotypes	<0.01	−0.13	—	0.12	0.04	0.28**	0.35**
Female-dominated occupations							
Knowledge	0.38**	0.11	−0.05	—	0.39***	−0.14	0.05
Interest	0.18	0.38***	−0.33**	0.27**	—	−0.08	0.14
Stereotypes	0.08	−0.11	0.54***	−0.09	−0.45***	—	0.27*
Felt pressure to conform	0.06	<0.01	−0.25*	−0.02	−0.22*	−0.03	—

Numbers represent the Pearson product moment correlation coefficients between designated scales. Below the main diagonal are results for boys, above are the results for girls. $^{*}p \leq .05$ $^{**}p \leq .01$ $^{***}p \leq .001$.

from 1 = only men to 7 = only women. Separate scores were averaged for male-dominated ($\omega = 0.70$; $\alpha = 0.68$) and female-dominated ($\omega = 0.70$; $\alpha = 0.70$) occupations, such that higher scores are consistent with the gender stereotype for the occupation. Each respective scale was four items each.

Felt Pressure. The measure was adapted from Patterson. (2012) revision of Egan and Perry. (2001) original scale (Patterson, 2012). In separate items, children rated on a 4-point scale (1 = really would not to 4 = really would) how they anticipated parents, teachers, and peers would respond if they engaged in gender non-conforming behaviors. Children indicated how likely each social agent would respond by 1) teasing, 2) being upset or unhappy, 3) trying to stop the behavior, and 4) trying to get them to act more like others of their own gender (total of 12 items; $\omega = 0.83$; $\alpha = 0.83$). A sample item from the girl's form is: "If you wanted to do something that boys usually do (but girls don't do), how much do you think other kids would try to stop you?" "Other kids" was replaced by parent and teacher to address the influence of each social Scores were the mean responses across the 12 items. Higher scores indicated greater perceived pressure.

RESULTS

Overview of Analysis

Data were analyzed in three phases that align with our aims. In the first phase, we explored whether there was a correlation between perceived knowledge and interest in occupations. In the second phase, we tested the hypothesis that there will be gender differences in occupational knowledge and interests using a series of independent samples *t-tests*. In the final step, we used hierarchical regressions to understand the combined predictive ability of perceived occupational knowledge, felt pressure, and gender stereotypes on occupational interests. Gender was dummy coded for all analyses (1 = girl; 0 = boy). All data analyses were carried out using SPSS statistical software and PROCESS macro for SPSS (Hayes, 2013).

Relation Between Perceived Knowledge and Interest

Correlations were calculated between perceived knowledge and interest for each set of occupations separately for boys and girls (**Table 1**). As expected, perceived knowledge was positively correlated with interest for both male- and female-dominated occupations for both boys and girls.

Gender Differences in Interest and Knowledge

It was expected that boys would have a greater interest in, and knowledge of male-dominated occupations compared to girls, and that the opposite would be true for female-dominated occupations. A series of independent samples *t-tests* confirmed the hypothesis (**Table 2**). Boys, compared to the girls, had more interest in and perceived knowledge of male-dominated occupations. Compared to boys, girls had more interest in and knowledge of female-dominated occupations. In addition, boys reported higher conformity pressure levels than girls. There were no gender differences in the stereotype measures.

Hierarchical Regression Analyses Testing Predictors of Occupational Interest

Hierarchical regression models examined the combined predictive ability of perceived occupational knowledge, felt pressure, and gender stereotypes on occupational interests. Additionally, models assessed if these effects differed for boys and girls. Separate models were calculated for interest in male- and female-dominated occupations. In each model, gender was entered in the first step. In the second step, perceived occupational knowledge, felt pressure, and gender stereotyping were entered. The interaction terms between gender and the primary variables were entered in the final step to test for moderation. The PROCESS macro (Hayes, 2013) was used to probe marginal and significant interaction terms. Data were mean centered. **Table 3** presents the standardized regression coefficients for each step of the models.

Predicting interest in female-dominated occupations (**Table 3**). We hypothesized that gender stereotypes and felt pressure would positively predict girls' interest, and negatively predict boys' interest, in female-dominated occupations. In the first step of the model, gender was a significant predictor, F (1, 157) = 63.71, $p < 0.001$, $R^2 = 0.29$. As expected, girls had a greater interest in female-dominated occupations than boys.

TABLE 2 | Mean gender differences in measures.

Measure	Boys M(SD)	Girls M(SD)	t	d
Perceived occupational knowledge				
Male-dominated occupations	2.95 (1.25)	2.18 (1.03)	4.49***	0.66
Female-dominated occupations	2.89 (1.34)	3.94 (0.95)	−5.98***	−0.87
Occupational interest				
Male-dominated occupations	2.31 (1.06)	1.65 (0.74)	4.85***	0.71
Female-dominated occupations	1.66 (0.81)	2.88 (1.02)	−8.87***	−1.33
Gender stereotyping				
Male-dominated occupations	5.33 (0.86)	5.18 (0.91)	1.16	—
Female-dominated occupations	5.31 (0.99)	5.36 (0.81)	−0.33	—
Gender socialization				
Felt pressure to conform	2.50 (0.63)	1.95 (0.64)	5.36***	0.85

Higher scores indicate greater perceived occupational knowledge, interest, stereotypes and felt pressure.
***p ≤ .001.*

TABLE 3 | Hierarchical linear regression model predicting occupational interest.

Dependent Variable/Predictors	Step 1	Step 2	Step 3
Female-Dominated Occupations[a]	**R^2 = 0.28*****	**ΔR^2 = 0.12*****	**ΔR^2 = 0.08*****
Gender	0.53***	0.43***	−1.22**
Felt pressure	—	−0.01	−0.20*
Knowledge	—	0.28***	0.17*
Gender stereotype	—	−0.18**	−0.33***
Pressure x gender	—	—	0.21*
Stereotype x gender	—	—	1.01**
Knowledge x gender	—	—	0.65**
Male-dominated occupations[b]	R^2 = 0.13***	ΔR^2 = 0.22***	ΔR^2 = 0.01
Gender	−0.37***	−0.24***	−0.97*
Felt Pressure	—	−0.02	−0.09
Knowledge	—	0.47***	0.40***
Gender Stereotype	—	−0.22*	−0.23**
Pressure x Gender	—	—	0.04
Stereotype x Gender	—	—	0.55
Knowledge x Gender	—	—	0.17

Entries are beta *coefficients. Participant Gender was coded dichotomously (0 = boy, 1 = girl).*
[a]*For the full model,* F *(7, 151) = 19.86,* p < *.001,* R^2 *= 0.48.*
[b]*For the full model,* F *(7, 152) = 8.34,* p < *.001,* R^2 *= 0.36.*
*p ≤ *.05,* **p ≤ *.01,* ***p ≤ *.001,*[m] p < *.08.*

There was a significant increase in variance explained from the first to the second block, $\Delta R^2 = 0.11$, ΔF (3, 154) = 9.60, $p < 0.001$. Gender remained a significant predictor. Perceived knowledge was positively associated with interest and stereotypes was negatively associated with interest. Felt pressure was not a significant predictor in the second step.

The addition of the interaction terms in the final step, resulted in a significant increase in variance explained, $\Delta R^2 = 0.08$, $p < 0.001$, ΔF (3, 151) = 7.50, $p < 0.001$. The interaction between felt pressure and gender was significant. Simple slopes analyses indicated that the effect of felt pressure on interest was marginally significant for boys (b = -0.29, $p = 0.05$) but not girls (b = 0.22, $p = 0.18$). The interaction between gender and stereotypes was also significant. Simple slopes analyses indicated that the effect of gender stereotypes on interest was significant for boys (b = -0.37, $p < 0.001$) but not girls (b = 0.10, $p = 0.39$).

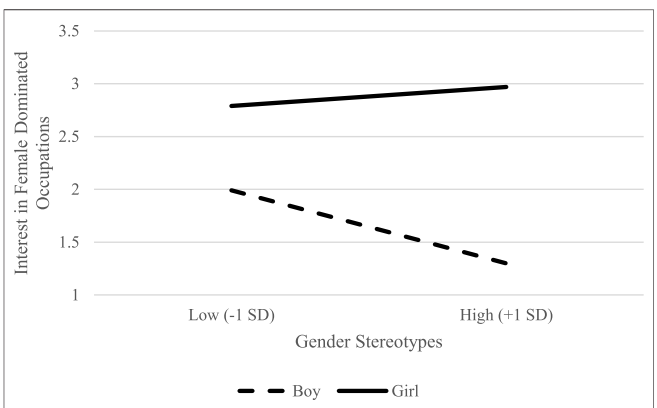


FIGURE 1 | Gender x gender stereotypes interaction.

FIGURE 2 | Gender x perceived occupational knowledge interaction.

Finally, the interaction between gender and perceived knowledge was significant for girls (b = 0.42, $p < 0.001$), and for boys (b = 0.17, $p = 0.01$). See **Figures 1**, **2**.

To summarize, these results show that the boys who felt less pressure to conform to gender norms were more interested in female-dominated occupations compared to boys who felt more pressure to conform. Additionally, boys were less interested in female-dominated occupations when they viewed them as predominately held by women. Finally, the effect of perceived knowledge on interest in female-dominated occupations was stronger for girls than boys.

Predicting interest in male-dominated occupations (**Table 3**). The first step of the model predicting interest in male-dominated occupations was significant, F (1, 157) = 24.57, $p < 0.001$, $R^2 = 0.13$, indicating that gender was a significant predictor. As hypothesized, boys were more interested in male-dominated occupations than girls. Results indicated a significant increase in variance explained from the first to the second block, $\Delta R^2 = 0.22$, ΔF (3, 154) = 17.48, $p < 0.001$. Gender remained a significant predictor. Although knowledge and stereotypes were significant, felt pressure was not. The increase in variance explained from the second block to the third block was not significant, $\Delta R^2 = 0.01$, ΔF (3, 151) = 1.00, $p > 0.05$.

To summarize, gender, perceived occupational knowledge and gender stereotypes were strong predictors of interest in male-dominated occupations. However, felt pressure did not appear to have an impact on interest in male-dominated occupations.

DISCUSSION

Gender differences in career interests that align with the segregated workforce are evident in children at early ages and remain relatively stable across the lifespan. As a result, these early career preferences may have a lasting impact and perpetuate the existing gender divide in work. Although researchers have extensively studied the impact of gender stereotypes (Oswald, 2008) on occupational interests, less work has examined how other facets of socialization, such as felt pressure to conform to gender norms and perceived occupational knowledge, may contribute to gender differences in children's occupational interest. The current study investigated whether gender, perceived occupational knowledge, occupational gender stereotypes, and as felt pressure to conform to gender norms predicted elementary aged children's occupational interests in a sample of low-SES children.

A significant contribution of this study is that it expands upon prior work and shows that perceived occupational knowledge predicts interest in gender-dominated occupations among fourth and fifth grade students. Although the observed relationship between perceived occupational knowledge and interest is not surprising, few studies have examined this relationship among younger children or children residing in the United States (Miller and Hayward, 2006). These results slightly differ from a study with older high school students in the United Kingdom that found that the relationship between perceived occupational knowledge and occupational interest was only significant for girls (Miller and Hayward, 2006). Perhaps there are developmental differences between elementary and high school boys' perceived occupational knowledge and occupational interests and the association between the two. Our hypothesis that perceived occupational knowledge is related to children's interest in gender-dominated was supported. These findings underscore how gender is embedded in everyday cognitive processes (Diekman and Schmader, 2020) and can result in gendered preferences even when there is no explicitly labeled categorization.

The gender differences observed in this sample confirmed the gender trends found in previous research (Ginevra and Nota, 2015). As hypothesized, boys rated their self-perceived knowledge and interest in male-dominated occupations higher than girls. Similarly, girls rated their self-perceived knowledge and interest in female-dominated occupations higher than boys. This finding is consistent with prior work demonstrating that youth are more interested in gender-typical occupations (Teig and Susskind, 2008; Coyle and Liben, 2018) and know more about occupations dominated by their own gender (Ferrari et al., 2015). Further, the finding that boys reported more felt pressure to conform to gender norms than girls replicates previous research (Egan and Perry, 2001; Smith and Leaper, 2006; Masters et al., 2020).

This study expands upon the current literature and demonstrates that pressure to conform to gender norms is related to children's occupational interests under some circumstances. Gender conformity pressure has been researched with regard to psychological adjustment among U.S. adolescents (e.g., Carver et al., 2003; Corby et al., 2007), educational motivations and efficacy (Leaper et al., 2012; Vantieghem et al., 2014) but not occupational interests. The effect of felt pressure to conform on interest in female-dominated occupations was moderated by gender. The more pressure boys felt to conform to gender norms, the less interested they were in female-dominated occupations. This finding is similar to other work that has highlighted that boys may avoid engaging in feminine-typed behaviors and characteristics (Halim and Ruble, 2010), rather than seeking to enact masculine-typed behaviors and characteristics.

Why was felt pressure to conform to gender norms related to boys', but not girls' occupational interests? Concurrently, children in middle childhood positively value conforming to gender norms (Egan and Perry, 2001) while expressing less rigidity in their gendered attitudes in some domains compared to younger children (Martin et al., 2002; Ruble et al., 2006). Given that boys face harsher social sanctions for violating gender norms compared to girls (Egan and Perry, 2001; Pauletti et al., 2017), it is plausible that boys are more attuned to external messages about what is *not* acceptable for them than messages about what is acceptable (Pauletti et al., 2017). Conversely, during childhood, girls may experience pressure to conform to feminine norms their parents in particular (Carr, 2007). Perhaps girls are more sensitive to messages about positive prescriptive stereotypes (desirable behaviors), rather than messages about negative proscriptive stereotypes (behaviors that one should avoid). The way in which felt pressure to conform was measured in this study may have captured pressure to *avoid* other-gender rather than pressure to *conform* to same-gender, normative behaviors. Further evidence and replication are required, specifically focusing on the circumstances under which felt pressure might affect boys and girls differently and distinguishing between prescriptive and proscriptive pressure.

In light of the findings for felt pressure, it is important to further consider why gender stereotypes influenced boys' disinterest in female-dominated occupations but did not significantly contribute to either boys' or girls' interest in male-dominated occupations. It is possible that these results differ in part due to the specific male-dominated occupations used in this study. The occupations examined are gender-segregated in the current workforce and represented as so in the media (Singh et al., 2020). However, male-dominated occupations that are commonly incorporated in other studies, such as doctors and pilots, are also associated with higher salaries than most female-dominated occupations. It is important for research on occupational gender stereotypes to not confound masculine and feminine occupations with differences in salaries (or education requirements) because such confounds cloud the interpretation of gender stereotype effects. Salary is associated with status, and indeed, all children are interested in occupations associated with high status (Teig and Susskind, 2008). In this study an attempt was made to choose masculine and feminine occupations that collectively were similar with respect to salary and education requirements. We view this as a strength of this research. Nevertheless, the strength of the relationship between boys' interest in male-dominated occupations and their perceptions of occupation gender stereotypes may be moderated by other factors such as salary and prestige.

One of the most important questions for future studies posed by these results is the pattern of relationships among the primary variables for girls. While the gendered nature of the occupations did seem to impact girls' interest in female-dominated occupations relative to boys, it is unclear why gender stereotyping or perceived pressure did not explain their interests. Work with older samples show that girls exhibit gender-typical occupational interests during childhood, but more gender balanced occupational interests emerge during adolescence (Sandberg et al., 1991; Helwig, 2008). Although girls were interested in female-dominated occupations more than boys, lower gender conformity pressure may be a precursor to less gendered interests later in adolescence. As described earlier, girls are allowed more flexibility than boys in their adherence to gender norms (Egan and Perry, 2001; Pauletti et al., 2017). Some evidence even suggests that girls are encouraged to engage with masculine domains. For example, a study of nine and ten-year-olds showed that girls who prefer male-typical activities are well-liked by their peers (Braun and Davidson, 2017). If masculine traits and activities are more socially valued than feminine traits and activities (Teig and Susskind, 2008), it makes sense that gender socialization and feminine stereotypes would have less of an impact girls' occupational interests. Alternatively, it is possible that these variables impact other constructs associated with male- and female-dominated occupational interests, such as self-efficacy, outcome expectations, and values. Indeed, prior work has shown that gender stereotypes (Brown, 2019) and gender socialization (Leaper et al., 2012) negatively impact girls' academic self-efficacy.

Limitations and Future Directions

The limitations of this study should be noted. First, the current study design does not allow for causal inferences and thus only provides an initial step in understanding the pathways between perceived occupational knowledge, felt pressure to conform to gender norms, and gender stereotyping, and gender differentiation of occupational interest. Second, it is possible that social desirability could have affected responses. For example, participants may have responded that occupations are appropriate for "both men and women" even though this response does not reflect their actual attitudes.

The occupations presented were selected carefully to avoid confounding gender stereotype designation with status related occupation characteristics, specifically salary. However, it is possible that these occupations may not reflect the breadth of potential occupations that children are currently interested in holding or may not represent the salient occupations held by men and women in these children's daily lives. As suggested by social role theory (Eagly, 1987) and social learning theory (Bussey and Bandura, 1999), it is important to consider the occupations of salient same-gender role models in these children's lives.

A strength of the present study was the focus on a study population of lower-SES children who have generally not been represented in this line of research. However, our findings may not generalize to children from middle- and high-SES backgrounds. For example, children from lower-SES backgrounds observe adults in different kinds of occupations than children from middle- or high-SES backgrounds. Indeed, our lower-SES sample may observe adults in occupational sectors with less gender equality more often than children from a higher-SES sample. Given that higher income adults are less likely than lower come adults to hold traditional gender beliefs (Katz-Wise et al., 2010), it is possible that children from lower-SES backgrounds internalize gender conformity pressure to a greater extent than those from middle- or high-SES

backgrounds. Future work should investigate the extent to which our findings parallel with a middle- or higher-SES sample.

Implications

This study makes theoretical contributions in the areas of gender and occupational interests. The finding that gender and perceived occupational knowledge were strong predictors of occupational interest lends support for gender development theories (Bem, 1981; Martin and Halverson, 1981) that suggest that children organize gendered information from their environment into schemas about what it means to be a girl or boy, which motivates their behaviors and may drive gender-typed personal preferences (Liben and Bigler, 2002; Ruble et al., 2006). The gender differences observed in perceived occupational knowledge and occupational interest expand upon social role theory (Eagly, 1987), which posits that gender differences in career interests result in part from the historical gender segregation in labor. Notably, this study may have detected an overlooked and understudied construct in research focused on gender differences in career interests, perceived occupational knowledge. For example, in our study, perceived knowledge differed for boys and girls, but stereotype beliefs did not. It will be important to consider how perceived knowledge is associated with constructs from models that are often used to predict gender differences in academic outcomes and career interests, such as self-competence or self-efficacy from expectancy-value theory (Eccles, 2011) and social cognitive career theory (Lent et al., 1994).

The findings have important implications related to workforce shortages due to gender segregation. The finding that perceived occupational knowledge is related to occupation gender composition by fourth grade suggests that interventions aimed at promoting children's interest in gender atypical occupations should be implemented at an earlier age. Such interventions are warranted especially when workforce shortages parallel the gender-segregated career choices (e.g., shortages in engineers, computer scientists, nurses and teachers). The gender differences highlighted point to the need for strategies to increase the perceived knowledge and interest in gender non-traditional occupations. To offset inferences that children make as a result of exposure to labor force segregation, it may be efficacious for career guidance counselors to directly address gender imbalances with children (Bigler and Liben, 2006). Although it is difficult to address the gender composition of occupations at the societal level, children may benefit from examples of individuals in gender non-traditional occupations (Dasgupta, 2011). Exposure to individuals in these occupations may be especially impactful for lower SES girls due to the plethora of low waged jobs typically done by women, such as housekeeping or food service, gender socialization may perpetuate poverty for these girls.

CONCLUSION

To better understand the reasons underlying gender differences in perceived occupational knowledge and interest, this study examined the role of gender stereotypes and gender socialization. This study confirms gender differences in career interests and knowledge and leads to new questions about factors that might determine these gender differences. It appears that boys and girls are equally cognizant of gender segregation in the workforce, but perhaps do not feel the same pressure to conform to these norms. The information gained from this study may be beneficial to further understand the interplay between perceived occupational knowledge, gender socialization, and gender stereotypes as they relate to children's occupational interests.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board at the University of Alabama. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SM and JB contributed to conception and design of the study. SM performed the statistical analysis. SM wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

REFERENCES

Antecol, H., and Cobb-Clark, D. A. (2013). Do psychosocial Traits Help Explain Gender Segregation in Young People's Occupations? *Labour Econ.* 21, 59–73. doi:10.1016/j.labeco.2012.12.005

Barth, J. M., Kim, H., Eno, C. A., and Guadagno, R. E. (2018). Matching Abilities to Careers for Others and Self: Do Gender Stereotypes Matter to Students in Advanced Math and Science Classes? *Sex Roles* 79 (1-2), 83–97. doi:10.1007/s11199-017-0857-5

Bem, S. L. (1981). Gender Schema Theory: A Cognitive Account of Sex Typing. *Psychol. Rev.* 88, 354–364. doi:10.1037/0033-295x.88.4.354

Bigler, R. S., and Liben, L. S. (2006). "A Developmental Intergroup Theory of Social Stereotypes and Prejudice," in *Advances in Child Development and Behavior*. Editor R. V. Kail (San Diego: Elsevier), 34, 39–89. doi:10.1016/S0065-2407(06)80004-2

Blakemore, J. E. O. (2003). Children's Beliefs about Violating Gender Norms: Boys Shouldn't Look like Girls, and Girls Shouldn't Act like Boys. *Sex Roles* 48, 411–419. doi:10.1023/A:1023574427720

Braun, S. S., and Davidson, A. J. (2017). Gender (Non)conformity in Middle Childhood: A Mixed Methods Approach to Understanding Gender-Typed Behavior, Friendship, and Peer Preference. *Sex Roles* 77, 16–29. doi:10.1007/s11199-016-0693-z

Brown, C. S. (2019). Sexualized Gender Stereotypes Predict Girls' Academic Self-Efficacy and Motivation across Middle School. *Int. J. Behav. Dev.* 43, 523–529. doi:10.1177/0165025419862361

Bureau of Labor Statistics (2016). Women in the Labor Force: A Data Book. Available at: https://www.bls.gov/opub/reports/womensdatabook/2016/home.htm.

Bussey, K., and Bandura, A. (1999). Social Cognitive Theory of Gender Development and Differentiation. *Psychol. Rev.* 106, 676–713. doi:10.1037/0033-295x.106.4.676

Carr, C. L. (2007). Where Have All the Tomboys Gone? Women's Accounts of Gender in Adolescence. *Sex Roles* 56, 439–448. doi:10.1007/s11199-007-9183-7

Carver, P. R., Yunger, J. L., and Perry, D. G. (2003). Gender Identity and Adjustment in Middle Childhood. *Sex Roles* 49, 95–109. doi:10.1023/a:1024423012063

Cook, R. E., Nielson, M. G., Martin, C. L., and Delay, D. (2019). Early Adolescent Gender Development: The Differential Effects of Felt Pressure from Parents, Peers, and the Self. *J. Youth Adolesc.* 48, 1912–1923. doi:10.1007/s10964-019-01122-y

Corby, B. C., Hodges, E. V., and Perry, D. G. (2007). Gender Identity and Adjustment in Black, Hispanic, and White Preadolescents. *Dev. Psychol.* 43, 261–266. doi:10.1037/00121649.43.1.26110.1037/0012-1649.43.1.261

Coyle, E. F., and Liben, L. S. (2018). Affecting Girls' Activity and Job Interests through Play: The Moderating Roles of Personal Gender Salience and Game Characteristics. *Child. Dev.* 87, 414–428. doi:10.1111/cdev.12463

Dasgupta, N. (2011). Ingroup Experts and Peers as Social Vaccines Who Inoculate the Self-Concept: The Stereotype Inoculation Model. *Psychol. Inq.* 22, 231–246. doi:10.1080/1047840x.2011.607313

Diekman, A., and Schmader, T. (2020). Gender as Embedded Social Cognition. doi:10.31234/osf.io/wvx2s

Doyle, E. (2011). Career Development Needs of Low Socio-Economic Status University Students. *Aust. J. Career Dev.* 20 (3), 56–65. doi:10.1177/103841621102000309

Eagly, A. H. (1987). *Sex Differences in Social Behavior: A Social Role Interpretation*. Hillsdale, NJ: Erlbaum.

Eccles, J. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *Int. J. Behav. Dev.* 35 (3), 195–201. doi:10.1177/0165025411398185

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). "Expectancies, Values, and Academic Behaviors," in *Achievement and Achievement Motivation*. Editor J. T. Spence (San Francisco, CA: W. H. Freeman), 75–146.

Eccles, J. S. (1987). Gender Roles and Women's Achievement-Related Decisions. *Psychol. Women Q.* 11 (2), 135–172. doi:10.1111/j.1471-6402.1987.tb00781.x

Eccles, J. S., Wigfield, A., Midgley, C., Reuman, D., Iver, D. M., and Feldlaufer, H. (1993). Negative Effects of Traditional Middle Schools on Students' Motivation. *Elem. Sch. J.* 93, 553–574. doi:10.1086/461740

Egan, S. K., and Perry, D. G. (2001). Gender Identity: A Multidimensional Analysis with Implications for Psychosocial Adjustment. *Dev. Psychol.* 37, 451–463. doi:10.1037//0012-1649.37.4.451

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behav. Res. Methods* 39, 175–191. doi:10.3758/BF03193146

Ferrari, L., Ginevra, M. C., Santilli, S., Nota, L., Sgaramella, T. M., and Soresi, S. (2015). Career Exploration and Occupational Knowledge in Italian Children. *Int. J. Educ. Vocat Guidance* 15 (2), 113–130. doi:10.1007/s10775-015-9299-1

Fornell, C., and Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *J. Marketing Res.* 18 (1), 39–50. doi:10.2307/3151312

Friedman, C. K., Leaper, C., and Bigler, R. S. (2007). Do Mothers' Gender-Related Attitudes or Comments Predict Young Children's Gender Beliefs? *Parenting* 7, 357–366. doi:10.1080/15295190701665656

Fulcher, M. (2011). Individual Differences in Children's Occupational Aspirations as a Function of Parental Traditionality. *Sex Roles* 64, 117–131. doi:10.1007/s11199-010-9854-7

Ginevra, M. C., and Nota, L. (2015). Occupational Gender Stereotypes and Problem-Solving in Italian Adolescents. *Br. J. Guidance Counselling* 45 (3), 312–327. doi:10.1080/03069885.2015.1063584

Gottfredson, L. S. (1981). Circumscription and Compromise: A Developmental Theory of Occupational Aspirations. *J. Couns. Psychol.* 28, 545–579. doi:10.1037/0022-0167.28.6.545

Haines, E. L., Deaux, K., and Lofaro, N. (2016). The Times They Are A-Changing . . . or Are They Not? A Comparison of Gender Stereotypes, 1983-2014. *Psychol. Women Q.* 40, 353–363. doi:10.1177/0361684316634081

Halim, M. L., Ruble, D., Tamis-LeMonda, C., and Shrout, P. E. (2013). Rigidity in Gender-Typed Behaviors in Early Childhood: A Longitudinal Study of Ethnic Minority Children. *Child. Dev.* 84, 1269–1284. doi:10.1111/cdev.12057

Halim, M. L., and Ruble, D. (2010). "Gender Identity and Stereotyping in Early and Middle Childhood," in *Handbook of Gender Research in Psychology*. Editors J. C. Christler and D. R. McCreary (New York: Springer), 495–525. doi:10.1007/978-1-4419-1465-1_24

Hartung, P. J. (2015). "Life Design in Childhood: Antecedents and Advancement," in *Life Design Handbook*. Editors L. Nota and J. Rossier (Göttingen: Hogrefe), 89–101.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression Based Approach*. New York: The Guilford Press.

Hayes, A. R., Bigler, R. S., and Weisgram, E. S. (2018). Of Men and Money: Characteristics of Occupations that Affect the Gender Differentiation of Children's Occupational Interests. *Sex Roles* 78 (11-12), 775–788. doi:10.1007/s11199-017-0846-8

Heinze, J. E., and Horn, S. S. (2014). Do Adolescents' Evaluations of Exclusion Differ Based on Gender Expression and Sexual Orientation? *J. Soc. Issues* 70 (1), 63–80. doi:10.1111/josi.12047

Helwig, A. A. (2008). From Childhood to Adulthood: A 15-year Longitudinal Career Development Study. *Career Dev. Q.* 57 (1), 38–50. doi:10.1002/j.2161-0045.2008.tb00164.x

Howard, K. A. S., and Walsh, M. E. (2011). Children's Conceptions of Career Choice and Attainment: Model Development. *J. Career Dev.* 38, 256–271. doi:10.1177/0894845310365851

Katz-Wise, S. L., Priess, H. A., and Hyde, J. S. (2010). Gender-role Attitudes and Behavior across the Transition to Parenthood. *Dev. Psychol.* 46, 18–28. doi:10.1037/a0017820

Koenig, A. M., and Eagly, A. H. (2014). Evidence for the Social Role Theory of Stereotype Content: Observations of Groups' Roles Shape Stereotypes. *J. Pers Soc. Psychol.* 107 (3), 371–392. doi:10.1037/a0037215

Leaper, C., Farkas, T., and Brown, C. S. (2012). Adolescent Girls' Experiences and Gender-Related Beliefs in Relation to Their Motivation in Math/science and English. *J. Youth Adolesc.* 41, 268–282. doi:10.1007/s10964-011-9693-z10.1007/s10964-011-9693-z

Lent, R. W., Brown, S. D., and Hackett, G. (1994). Toward a Unifying Social Cognitive Theory of Career and Academic Interest, Choice, and Performance. *J. Vocational Behav.* 45, 79–122. doi:10.1006/jvbe.1994.1027

Liben, L. S., and Bigler, R. S. (2002). The Developmental Course of Gender Differentiation: Conceptualizing, Measuring, and Evaluating Constructs and Pathways. *Monogr. Soc. Res. Child. Dev.* 67 (2), 147. doi:10.1111/1540-5834.t01-1-00187

Low, K. S. D., Yoon, M., Roberts, B. W., and Rounds, J. (2005). The Stability of Vocational Interests from Early Adolescence to Middle Adulthood: a Quantitative Review of Longitudinal Studies. *Psychol. Bull.* 131, 713–737. doi:10.1037/0033-2909.131.5.713

Martin, C. L., Ruble, D. N., and Szkrybalo, J. (2002). Cognitive Theories of Early Gender Development. *Psychol. Bull.* 128, 903–933. doi:10.1037/0033-2909.128.6.903

Martin, C. L., and Halverson, C. F., Jr. (1981). A Schematic Processing Model of Sex Typing and Stereotyping in Children. *Child. Dev.* 52 (4), 1119–1134. doi:10.2307/1129498

Masters, S. L., Hixson, K., and Hayes, A. R. (2020). Perceptions of Gender Norm Violations Among Middle School Students: An Experimental Study of the Effects of Violation Type on Exclusion Expectations. *The J. Early Adolescence* 41, 527–549. doi:10.1177/0272431620931193

McMahon, M., and Patton, W. (1997). Gender Differences in Children and Adolescents' Perceptions of Influences on Their Career Development. *Sch. Counselor* 44, 368–376.

Miller, L., and Hayward, R. (2006). New Jobs, Old Occupational Stereotypes: Gender and Jobs in the New Economy. *J. Edu. Work* 19 (1), 67–93. doi:10.1080/13639080500523000

Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*. 3rd ed. New York, NY: McGraw-Hill.

Oswald, D. L. (2008). Gender Stereotypes and Women's Reports of Liking and Ability in Traditionally Masculine and Feminine Occupations. *Psychol. Women Q.* 32, 196–203. doi:10.1111/j.1471-6402.2008.00424.x

Pacilli, M. G., Spaccatini, F., Barresi, C., and Tomasetto, C. (2019). Less Human and Help-Worthy: Sexualization Affects Children's Perceptions of and Intentions toward Bullied Peers. *Int. J. Behav. Dev.* 43, 481–491. doi:10.1177/0165025419873040

Pascoe, C. (2012). "Dude, You're a Fag: Masculinity and Sexuality in High School," in *EBooks on Demand)*. 2nd ed. (Berkeley: University of California Press).

Patterson, M. (2012). Self-perceived Gender Typicality, Gender-Typed Attributes, and Gender Stereotype Endorsement in Elementary-School-Aged Children. *Sex Roles* 67 (7–8), 422–434. doi:10.1007/s11199-012-0184-9

Pauletti, R. E., Menon, M., Cooper, P. J., Aults, C. D., and Perry, D. G. (2017). Psychological Androgyny and Children's Mental Health: A New Look with New Measures. *Sex Roles* 76, 705–718. doi:10.1007/s11199-016-0627-9

Pedhazur, E. J., and Schmelkin, L. P. (2013). *Measurement, Design, and Analysis: An Integrated Approach*. New York: Psychology Press.

Rice, L., Barth, J. M., Guadagno, R. E., Smith, G. P., and McCallum, D. M. (2013). The Role of Social Support in Students' Perceived Abilities and Attitudes toward Math and Science. *J. Youth Adolesc.* 42, 1028–1040. doi:10.1007/s10964-012-9801-8

Rohlfing, J. E., Nota, L., Ferrari, L., Soresi, S., and Tracey, T. J. G. (2012). Relation of Occupational Knowledge to Career Interests and Competence Perceptions in Italian Children. *J. Vocational Behav.* 81, 330–337. doi:10.1016/j.jvb.2012.08.001

Ruble, D. N., Martin, C. L., and Berenbaum, S. A. (2006). "Gender Development," in *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality Development*. 6th ed., Series Editors W. Damon, R. M. Lerner, and N. Eisenberg (Hoboken, NJ: Wiley), 858–932.

Sandberg, D. E., Ehrhardt, A. A., Ince, S. E., and Meyer-Bahlburg, H. F. L. (1991). Gender Differences in Children's and Adolescents' Career Aspirations: A Follow-Up Study. *J. Adolesc. Res.* 6, 371–386. doi:10.1177/074355489163007

Schmitt-Wilson, S., and Welsh, M. C. (2012). Vocational Knowledge in Rural Children: A Study of Individual Differences and Predictors of Occupational Aspirations and Expectations. *Learn. Individual Differences* 22 (6), 862–867. doi:10.1016/j.lindif.2012.06.003

Singh, V. K., Chayko, M., Inamdar, R., and Floegel, D. (2020). Gender Bias in Occupational Images on Digital media Platforms. *J. Assoc. Inf. Sci. Technol.* 71, 1281–1294. doi:10.1002/asi.24335

Smith, T. E., and Leaper, C. (2006). Self-perceived Gender Typicality and the Peer Context during Adolescence. *J. Res. Adolescence* 16, 91–104. doi:10.1111/j.1532-7795.2006.00123.x

Spence, J. T., and Hall, S. K. (1996). Children's Gender-Related Self-Perceptions, Activity Preferences, and Occupational Stereotypes: A Test of Three Models of Gender Constructs. *Sex Roles* 35, 659–691. doi:10.1007/bf01544086

Stockard, J., and McGee, J. (1990). Children's Occupational Preferences: The Influence of Sex and Perceptions of Occupational Characteristics. *J. Vocational Behav.* 36, 287–303. doi:10.1016/0001-8791(90)90033-X

Sullivan, J., Moss-Racusin, C., Lopez, M., and Williams, K. (2018). Backlash against Gender Stereotype-Violating Preschool Children. *Plos ONE* 13, e0195503–24. doi:10.1371/journal.pone.0195503

Teig, S., and Susskind, J. E. (2008). Truck Driver or Nurse? the Impact of Gender Roles and Occupational Status on Children's Occupational Preferences. *Sex Roles* 58, 848–863. doi:10.1007/s11199-008-9410-x

Vantieghem, W., Vermeersch, H., and van Houtte, M. (2014). Transcending the Gender Dichotomy in Educational Gender gap Research: The Association between Gender Identity and Academic Self-Efficacy. *Contemp. Educ. Psychol.* 39, 369–378. doi:10.1016/j.cedpsych.2014.10.001

Watson, M., and McMahon, M. (2005). Children's Career Development: A Research Review from a Learning Perspective. *J. Vocational Behav.* 67, 119–132. doi:10.1016/j.jvb.2004.08.011

Weinger, S. (2000). Economic Status: Middle Class and Poor Children's Views. *Child. Soc.* 14, 135–146. doi:10.1002/(sici)1099-0860(200004)14:2<135:aid-chi592>3.3.co;2-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 24 March 2022
doi: 10.3389/fpsyg.2022.818607

Gender Differences in the Associations Between Perceived Parenting Styles and Young Adults' Cyber Dating Abuse

F. Giorgia Paleari, Laura Celsi, Desirèe Galati and Monica Pivetti*

Department of Human and Social Sciences, University of Bergamo, Bergamo, Italy

Existing literature indicates that parenting styles affect the development of cyber aggression in offspring differently, depending on the gender of children. The present study investigates whether mothers' and fathers' parenting styles show similar gender differences in their associations with a new form of dating violence, i.e., cyber dating abuse (CDA). The limited evidence on the issue focuses on the relation that each parenting style has with CDA perpetration, without considering CDA victimization and the joint effects of fathers' and mothers' parenting styles. The present study contributes to the research on gender differences in parenting by examining whether young adults' perceptions of maternal and paternal parenting styles during childhood were independently and/or jointly related to their perpetrated and suffered CDA and whether these relations differed across young adults' gender. In total, 351 young adults (50.7% men), age between 18 and 35 years and having a romantic relationship, completed online self-reports of the variables of interest that include a bidimensional measure of perpetrated/suffered CDA that assess aggression and control. Results showed that maternal authoritarian parenting was uniquely and positively associated to their children's perpetration and victimization of cyber dating control, whereas maternal permissive parenting was uniquely and positively related to their children's perpetration of cyber dating aggression and victimization of cyber dating control. For daughters, these associations were stronger when the father's style was similar to the mother's one or when a maternal authoritarian style combined with a paternal permissive style, thus indicating that the two parents' parenting styles interact in relating to their daughters' CDA.

Keywords: parenting styles, cyber dating abuse, gender differences, young adults, gendered socialization

OPEN ACCESS

Edited by:
Joyce Endendijk,
Utrecht University, Netherlands

Reviewed by:
Maria Muñiz,
Universidad Pablo de Olavide, Spain
Sónia Caridade,
University of Minho, Portugal

****Correspondence:***
F. Giorgia Paleari
francesca-giorgia.paleari@unibg.it

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *19 November 2021*
Accepted: *10 February 2022*
Published: *24 March 2022*

Citation:
Paleari FG, Celsi L, Galati D and Pivetti M (2022) Gender Differences in the Associations Between Perceived Parenting Styles and Young Adults' Cyber Dating Abuse. Front. Psychol. 13:818607. doi: 10.3389/fpsyg.2022.818607

INTRODUCTION

The family of origin usually is the most important socialization agent in the early stages of individuals' development. Several theoretical models, such as social learning theory (Bandura, 1977), coercion theory (Reid et al., 2002), attachment theory (Michiels et al., 2008), and the self-determination theory (Soenens et al., 2015), suggest that parents may affect children's

and adolescents' behaviors and their peer relationships through their parenting choices, practices, and beliefs.

According to Baumrind (1971, 1991), the construct that best summarizes the main factors through which parents influence the socio-emotional and behavioral development of their children is parenting style, that is the pervasive emotional climate within which the child is raised (Darling and Steinberg, 1993). Baumrind 's (1971, 1991) categorical model distinguishes three different parenting styles–authoritative, authoritarian, and permissive–on the basis of four parental behavior dimensions: parental warmth, control, demand, and involvement. More specifically, authoritative style, characterized by high levels of control, demand, parental warmth, and involvement, is recognizable in affective and sensitive parents, who discipline their children through open communication and example, have high but reasonable demands, and are strict but fair. The authoritarian style, characterized by high levels of control and demand and low levels of warmth and involvement, is evident in strict and inflexible parents, who show high expectations toward their children, little sensitivity toward emotional needs of children, and punish them without explaining the meaning of the rules imposed. Finally, the permissive style, characterized by high levels of parental warmth and involvement and low levels of control and demand, is identifiable in caring, affective, and sensitive parents, who exercise the role of friends rather than parents and thus display an excessive indulgence and poor ability in the exercise of normative functions.

The wide empirical literature inspired by this model generally attests that the authoritarian and, to a lesser degree, the permissive parenting style contribute to the development of behavioral problems, such as the perpetration of bullying and dating violence in offspring (Luk et al., 2016; Olivari et al., 2017; Pinquart, 2017; Cuccì et al., 2019; Ruiz-Hernández et al., 2019; Moreno Méndez et al., 2020). Conversely, the authoritative parenting has a protective effect against externalizing problems and both perpetration and victimization of relationship abuse, even in the presence of parental inconsistency (Luk et al., 2016; Mumford et al., 2016; Pinquart, 2017; Ruiz-Hernández et al., 2019). These effects resulted not moderated by child and parent gender (Pinquart, 2017). The cultural invariance of the above findings was, however, questioned by recent research in Latin American and Mediterranean European countries, where permissive parenting was found to have more positive outcomes than expected (Martínez et al., 2019; Suárez-Relinque et al., 2019).

The digital revolution has caused such substantial changes within relational dynamics, especially among current adolescents and young adults belonging to the Y and the Z generations (Buckingham and Willett, 2006; Junco and Mastrodicasa, 2007), that scholars have been forced to rethink the construct of violence in a way that also includes the virtual world. Recent studies have indeed highlighted the rapid spread in cyber space of new forms of intentional acts harming individuals or groups, which has been given the name of cyber aggression (Zhao and Gao, 2012; Zhang et al., 2021). Results available to date on the role of parenting styles in predicting offspring cyber aggression are only partially consistent with those concerning violence in the real world. In fact, several studies show that the parental authoritarian style positively relates to children's perpetration and victimization of cyber aggression; however, the relation between authoritarian parenting and perpetrated cyber aggression relation seems stronger for men than for women, suggesting that the authoritarian style fosters greater assimilation of traditional gender roles in which violence is less criticized in boys (Elsaesser et al., 2017; He et al., 2017; Martínez-Ferrer et al., 2019; Moreno-Ruiz et al., 2019; Zhang et al., 2021). In addition, results linking cyber aggression to the other two parenting styles seem more inconsistent: some reveal that parental indulgent and authoritative styles relate negatively with cyber violence, whereas some others indicate they are unrelated or positively related to it (e.g., Vale et al., 2018; Moreno-Ruiz et al., 2019; Zhang et al., 2021).

Among the various forms of cyber aggression, cyber dating abuse (CDA) refers to acts of control, aggression, and sexual coercion that are digitally perpetrated against the romantic partner through new media, such as social network sites, text messages, emails, or technology, such as geolocation app (Zweig et al., 2013, 2014; Borrajo et al., 2015; Reed et al., 2017). CDA appears to be widespread and dangerous for the mental health of both victims and perpetrators, resulting in externalizing and internalizing symptoms (Drauker and Martsolf, 2010; Bennet et al., 2011; Zweig et al., 2014; Sargent et al., 2016; Flach and Deslandes, 2017; Van Ouytsel et al., 2017).

Regarding CDA etiology, some evidence suggests that adverse childhood experiences lived in the family, such as experiencing abuse and witnessing intimate partner violence (IPV), are related to an increased likelihood of CDA perpetration and victimization, directly or through the internalization of early maladaptive relational schemas (Celsi et al., 2021; Smith-Darden et al., 2016; Ramos et al., 2017). However, not much attention has been devoted to other family of origin factors that may contribute to CDA. Particularly, only one study by Muñiz-Rivas et al. (2019) has recently examined which parenting style best predicts the risk of CDA perpetration. Their findings indicate that male and female adolescents with authoritarian mothers were the most prone to inflict cyber dating aggression and cyber dating control, respectively, whereas adolescents with indulgent mothers were the less prone. The authors explained the greater influence of mothers' parenting styles as the consequence of their greater involvement in daily child-rearing, especially in domains related to affective relationships. Indeed, mothers are expected to be and remain the main caregiver despite a steady increase in women's participation in work outside of the home (Raley et al., 2012). Muñiz-Rivas et al. (2019), however, omit to assess CDA victimization and cyber sexual coercion and do not examine the joint effects of fathers and mothers' parenting styles, despite there is evidence that the combination of the two parents' styles can explain more variance in children's externalizing behaviors than the focus on only one parent's style (Berkien et al., 2012).

Informed by the literature just reviewed, the present research aimed at investigating whether young adults' perceptions of maternal and paternal parenting styles during childhood were

independently and/or jointly related to their perpetrated and suffered CDA, with focused attention on gender differences.

As for the unique relations of parenting styles with CDA, we hypothesized that independently of witnessing IPV between parents, the more young adults reported their mother or father as having been authoritarian, the more they perpetrated and suffered CDA (H1); authoritarian parenting was more strongly related to young adults' perpetrated CDA when mothers', rather than fathers', parenting was considered (H2) and in men, rather than in women (H3). We were unable to make well-founded predictions about the association of permissive and authoritative styles with CDA, because of previous studies conflicting results relating those styles to cyber aggression and to CDA. Similarly, no specific predictions were made about the joint relations of parenting styles with CDA due to the lack of evidence on the issue.

METHODS

Participants and Procedure

Participants were 351 young adults, 49.3% were women and 50.7% were men, aging on average 24 years ($M = 24.20$; $SD = 3.20$; range: 18–35). Their most frequent education qualifications were high school diploma or equivalent (46.4%), bachelor degree (28.8%), and master degree (21.1%).

All of them were engaged in a romantic relationship, mainly a heterosexual one (96.6%), averaging 3.62 years ($SD = 2.99$; range: 1 month–24 years). Most participants (79.2%) were not cohabiting with their romantic partners. All subjects had grown up with their parents.

On average, participants referred to use smartphones very often ($M = 6.00$; $SD = 1.07$) and social networks often ($M = 5.24$; $SD = 1.36$; possible range of response for both variables: from 1 = never to 7 = always).

Men and women did not differ with respect to any of the above socio-demographics except for social networks use, which was more frequent for women ($M = 5.47$) than for men [$M = 5.02$; $t(349) = 3.147$, $p = 0.002$].

Subjects were contacted through the publication of a post on instant messaging platforms, which presented the study as an anonymous survey on family and couple relationships and specified the inclusion criteria (identifying oneself as male or female, aging between 18 and 35 years, and having a romantic relationship lasting for at least 1 month). The message also contained a link to the online survey and asked participants to disseminate it to acquaintances. Informed consent was obtained from participants. The study complied with the Ethics Code of the Italian Psychology Association (Associazione Italiana di Psicologia [AIP], 2015) and was conducted in accordance with the (World Medical Association, 2013)-Declaration of Helsinki (1964/2013).

Measures

Parenting Styles

Young adults' perceptions of their parents' parenting practices during childhood were measured through the 40-item Italian version of the Parenting Styles and Dimensions Questionnaire (PSDQ; Tagliabue et al., 2014). The participant responded to two versions of the scale, one for the mother's parenting style and one for the father's. The scale assesses the three parenting styles suggested by Baumrind (1971, 1991): authoritative (23 items, e.g., "My mother/father encouraged me to talk about my troubles"; $\alpha = 0.98$ for both mothers and fathers) (see **Supplementary Material** for internal consistencies for men and women, separately), authoritarian (13 items, e.g., "My mother/father guided me by punishment more than by reason"; $\alpha = 0.92$ and 0.94 for mothers and fathers, respectively), and permissive (4 items, e.g., "My mother/father stated punishments to me and did not actually did them"). Since the permissive subscale had shown low reliability in previous studies (e.g., Tagliabue et al., 2014), we increased it by adding to the subscale 10 more items from the original version of the PSDQ (Robinson et al., 2001) ($\alpha = 0.80$ and 0.77 for mothers and fathers, respectively).

Perpetrated and Suffered Cyber Dating Abuse

Perpetrated and suffered CDA within the current romantic relationship was measured through a scale previously validated in Italy by Celsi et al. (2021). The scale consists of 40 items (20 for perpetration and 20 for victimization) assessing two dimensions of CDA: monitoring and control (11 items, e.g., "I/my partner checked my/my partner's location and online activities"; $\alpha = 0.86$ and 0.89 for perpetration and victimization, respectively) and psychological or sexual pressure and aggression (9 items, e.g., "I/my partner sent a threatening message to my partner/me"; $\alpha = 0.84$ and 0.78 for perpetration and victimization, respectively).

Intimate Partner Violence Perpetrated by Parents

Physical and psychological IPV perpetrated by parents and witnessed by respondents during their childhood was assessed through a 6-item measure by Celsi et al. (2021). Three items measured violence perpetrated by the mother against the father and three items assessed violence perpetrated by the father against the mother (e.g., "I saw/heard my mother/father being insulted, denigrated, humiliated, or verbally assaulted by my father/mother"; $\alpha = 0.71$ and 0.84 for violence perpetrated by mothers and fathers, respectively).

Participants responded to the items of the three measures using a 7-point Likert scale ranging from 1 (never) to 7 (always).

Data Analysis

Hypotheses were verified using multiple regression analyses in SPSS, combined with Hayes' (2013) PROCESS macros for Model 1, testing simple moderations (or 2-wave interactions), and Model 3, testing moderated moderations (or 3-wave interactions) (for more details see **Supplementary Material**). All PROCESS analyses were performed controlling for the parenting styles others the ones entered as the predictor and the moderator and for father and mother perpetrated IPV and child networks use (which resulted to differ across gender).

In order to address non-normality that is common in CDA and IPV data, the bootstrap technique ($N = 5{,}000$) was used to compute CIs.

RESULTS

Preliminary Results

When compared to women, on average men reported that their mother had been more permissive ($M = 2.58$ and 2.41; t-test (349) = 2.062, 95% CI [0.01; 0.33]) and their father had perpetrated less IPV ($M = 1.46$ and 1.87; t-test (349) = −3.148, 95% CI [−0.64; −0.17]). As concerns CDA, men resulted to perpetrate more aggression ($M = 1.22$ and 1.10; t-test (349) = 2.855, 95% CI [0.04; 20]) and less control ($M = 1.57$ and 1.82; t-test (349) = −3.034, 95% CI [−0.42; −0.09]) and to suffer more control ($M = 1.70$ and 1.45; t-test (349) = 2.924, 95% CI [0.08; 0.43]) and more aggression ($M = 1.22$ and 1.12; t-test (349) = 2.611, 95% CI [0.03; 0.18]) than women did (see **Supplementary Table 2** for descriptive statistics and correlations).

Unique Relations of Parenting Styles With Cyber Dating Abuse

Regression models indicated that mother but not father parenting styles were uniquely but weakly related to their child CDA (see **Table 1**). In particular, the more the mother was perceived as authoritarian the more the child perpetrated and suffered cyber dating control; also, the more the mother was judged as permissive the more the child perpetrated cyber dating aggression and suffered cyber dating control. PROCESS Model 1 revealed that none of the unique associations between parenting styles and CDA was moderated by the gender of participants.

Joint Relations of Parenting Styles With Cyber Dating Abuse

From PROCESS Model 1, we found that only the mother authoritarian style and the father permissive style interacted in relating to their children CDA. Specifically, the association of mother authoritarian style with both perpetrated and suffered cyber dating aggression was stronger the more permissive the father was (2-wave interaction effects: $B = 0.06$, β = 0.13, 95% CI [01, 0.10], $f^2 = 0.02$[1] and $B = 0.05$, β = 0.14, 95% CI [0.01, 0.10], $f^2 = 0.02$ for perpetrated and suffered aggression, respectively). Simple slope tests showed that such associations were significant only for children having a more permissive father (1 SD above the mean) ($B = 0.06$, β = 0.18, 95% CI [0.01, 0.11] and $B = 0.06$, β = 0.21, 95% CI [0.01, 0.11] for perpetrated and suffered aggression, respectively).

Finally, PROCESS Model 3 showed that mother and father parenting styles interacted in relating to CDA differently for daughters and sons. Specifically, the previous interaction effects

[1] f^2 effect sizes of 0.02, 0.15, and 0.35 can be considered small, medium, and large, respectively (Cohen, 1988). However, when interpreting the interaction term's impact, these cutoffs result overly conservative and are commonly replaced by f^2 values of 0.005, 0.01, and 0.025 (Aguinis et al., 2005; Kenny, 2018).

TABLE 1 | The role of parenting styles on CDA when controlling for networks use and IPV perpetrated by parents.

	Perpetrated CDA–control				Perpetrated CDA–aggression				Suffered CDA–control				Suffered CDA–aggression			
	$R^2 = 0.09$, $p = 0.000$ Cohen's $f^2 = 0.10$				$R^2 = 0.07$, $p = 0.005$ Cohen's $f^2 = 0.08$				$R^2 = 0.06$. $p = 0.017$ Cohen's $f^2 = 0.06$				$R^2 = 0.06$, $p = 0.016$ Cohen's $f^2 = 0.06$			
	B	SE	β	95% CI	B	SE	β	95% CI	B	SE	B	95% CI	B	SE	β	95% CI
M Authoritative PS	−0.01	0.04	−0.02	[−0.08; 0.05]	0.01	0.02	0.04	[−0.02; 0.05]	0.01	0.04	0.03	[−0.07; 0.10]	0.00	0.03	0.01	[−0.05; 0.05]
M Authoritarian PS	0.09	0.05	**0.14**	[0.01; 0.18]	0.02	0.02	0.07	[−0.02; 0.07]	0.11	0.05	**0.17**	[0.01; 0.23]	0.03	0.03	0.09	[−0.02; 0.09]
M Permissive PS	−0.01	0.07	−0.01	[−0.14; 0.13]	0.06	0.03	**0.13**	[0.01; 0.14]	0.16	0.07	**0.15**	[0.01; 0.36]	0.05	0.04	0.11	[−0.01; 0.13]
F Authoritative PS	0.03	0.03	−0.01	[−0.0; 3.09]	0.01	0.02	0.05	[−0.01; 0.04]	−0.02	0.04	−0.06	[−0.10; 0.06]	0.00	0.02	−0.01	[−0.04; 0.04]
F Authoritarian PS	−0.02	0.04	0.06	[−0.10; 0.07]	0.00	0.02	−0.01	[−0.04; 0.04]	−0.03	0.04	−0.05	[−0.14; 0.08]	0.01	0.02	0.04	[−0.04; 0.05]
F Permissive PS	0.09	0.07	−0.03	[−0.05; 0.24]	0.00	0.03	−0.01	[−0.06; 0.06]	−0.03	0.07	−0.03	[−0.17; 0.10]	0.01	0.04	0.02	[−0.06; 0.08]
IPV perpetrated by M	0.09	0.06	0.09	[−0.06; 0.24]	0.08	0.03	**0.18**	[0.02; 0.18]	−0.05	0.06	−0.05	[−0.17; 0.08]	0.04	0.03	0.10	[−0.01; 0.11]
IPV perpetrated by F	0.07	0.05	0.10	[−0.06; 0.21]	0.01	0.02	0.03	[−0.07; 0.10]	0.04	0.05	0.05	[−0.08; 0.17]	0.00	0.03	−0.01	[−0.06; 0.06]
Networks use	0.08	0.03	**0.14**	[0.03; 0.14]	0.01	0.02	0.03	[−0.02; 0.04]	0.05	0.05	0.08	[−0.01; 0.11]	0.00	0.02	0.05	[−0.02; 0.05]

M, mother; F, father; PS, parenting style; IPV, intimate partner violence; CDA, cyber dating abuse.
Significant results are typed in bold.

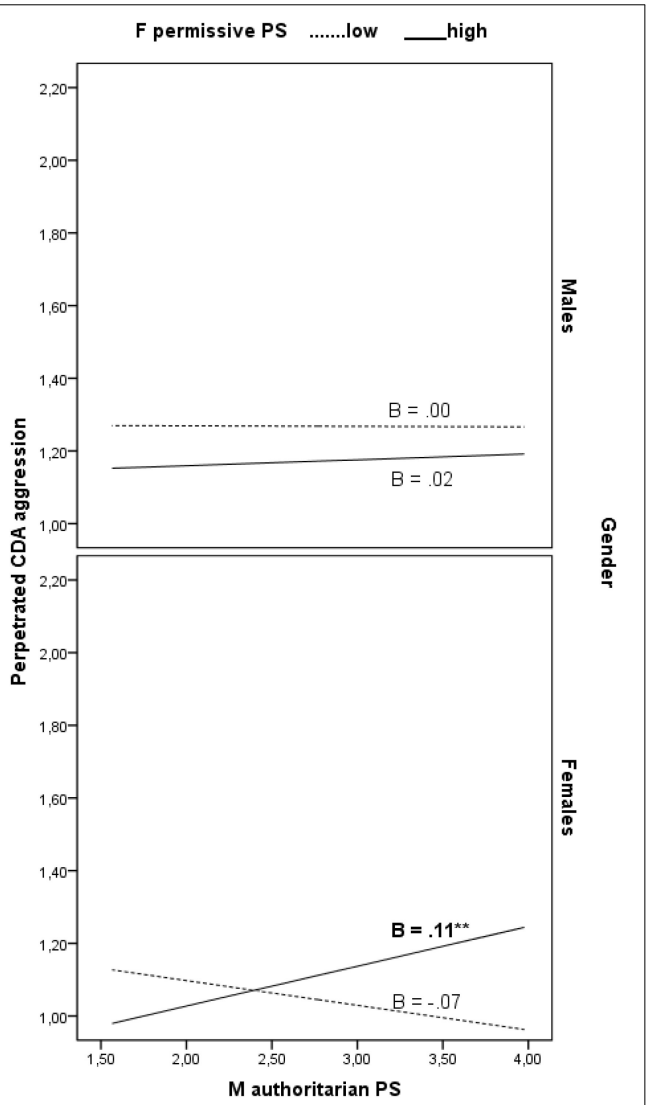


FIGURE 1 | The conditional effects of mother authoritarian parenting style on perpetrated cyber dating aggression as a function of father permissive parenting style and child gender. $^{**}p < 0.01$. M, mother; F, father; PS, parenting style; CDA, cyber dating abuse.

of mother authoritarian style and father permissive style on perpetrated cyber dating aggression were significantly moderated by child gender (3-wave interaction effects: $B = 0.11$, $\beta = 0.13$, 95% CI [0.02, 0.19], $f^2 = 0.02$). Simple slope test showed that mother authoritarian style was significantly associated with a higher degree of perpetrated cyber dating aggression only in daughters having a more permissive father (1 SD above the mean) ($B = 0.11$, $\beta = 0.34$, 95% CI [0.04, 0.18]) (see **Figure 1**).

In addition, the association of mother authoritarian style with perpetrated and suffered cyber dating control varied as a function of both father authoritarian style and child gender (3-wave interaction effects: $B = 0.10$, $\beta = 0.10$, 95% CI [01, 0.19], $f^2 = 0.01$ and $B = 0.10$, $\beta = 0.11$, 95% CI [01, 0.20], $f^2 = 0.01$ for perpetrated and suffered control, respectively). Simple slope tests showed that mother authoritarian style was significantly associated with higher degrees of perpetrated and suffered cyber dating control in daughters having more authoritarian fathers (1 SD above the mean) ($B = 0.17$, $\beta = 0.25$, 95% CI [0.03, 0.30] and $B = 0.16$, $\beta = 0.10$, 95% CI [0.02, 0.30] for perpetrated and suffered control, respectively). In addition, mother authoritarian style was significantly associated with a higher degree of suffered cyber dating control in sons having poorly authoritarian fathers (1 SD below the mean) ($B = 0.19$, $\beta = 0.28$, 95% CI [0.03, 0.35]) (see **Figure 2**).

Finally, mother permissive style was differently related to suffered cyber dating aggression and control as a function of father permissive style and child gender (3-wave interaction effects: $B = 0.11$, $\beta = 0.09$, 95% CI [0.01, 0.22], $f^2 = 0.01$ and $B = 0.46$, $\beta = 0.17$, 95% CI [0.22, 0.70], $f^2 = 0.04$ for suffered aggression and suffered control, respectively). Simple slope tests showed that mother permissive style was significantly associated with higher degrees of suffered cyber dating aggression and control only in daughters having more permissive fathers (1 SD above the mean) ($B = 0.10$, $\beta = 0.22$, 95% CI [0.01, 0.21] and $B = 0.45$, $\beta = 0.42$, 95% CI [0.22, 0.68] for suffered aggression and suffered control, respectively) (see **Figure 3**).

DISCUSSION

Our results showed that the more young adults reported that their mothers had been authoritarian or permissive during their childhood the more likely they were to be involved in a cyber abusive dating relationship. In fact, when controlling for the confounding effects of IPV and networks use, mothers' authoritarian parenting was uniquely, albeit weakly, associated to their children's perpetration and victimization of cyber dating control. Partially in line with our prediction (H1), these results support the expected relation between authoritarian parenting and CDA, but only when mothers' parenting and the control dimension of CDA were considered. Thus, young adults raised by more authoritarian mothers (who were coercive and controlling, but poorly empathic and warm) tend to replicate and bear controlling practices when interacting online with their partners. According to social learning theory (Bandura, 1977), children who are exposed to controlling parents may view their parents' behaviors as acceptable or desirable and model their interpersonal behaviors based on them, therefore engaging more controlling behaviors with their partner and tolerating more controlling behaviors by him/her (Curry and Zavala, 2020). Alternatively, attachment theory (Bowlby, 1969) posits that coercive family processes facilitate the development of insecure attachment, which in turn contributes to personality characteristics, such as separation anxiety, partner jealousy, and distrust, which likely increase partner surveillance (Guerrero, 1998; Mikulincer and Shaver, 2010; Buck et al., 2012).

Moreover, mothers' permissive parenting was uniquely, albeit weakly, associated to their children's perpetration of cyber dating aggression and victimization of cyber dating control. Young adults raised by permissive parents are less used to be controlled and, because of the few guidelines and limited rules received,

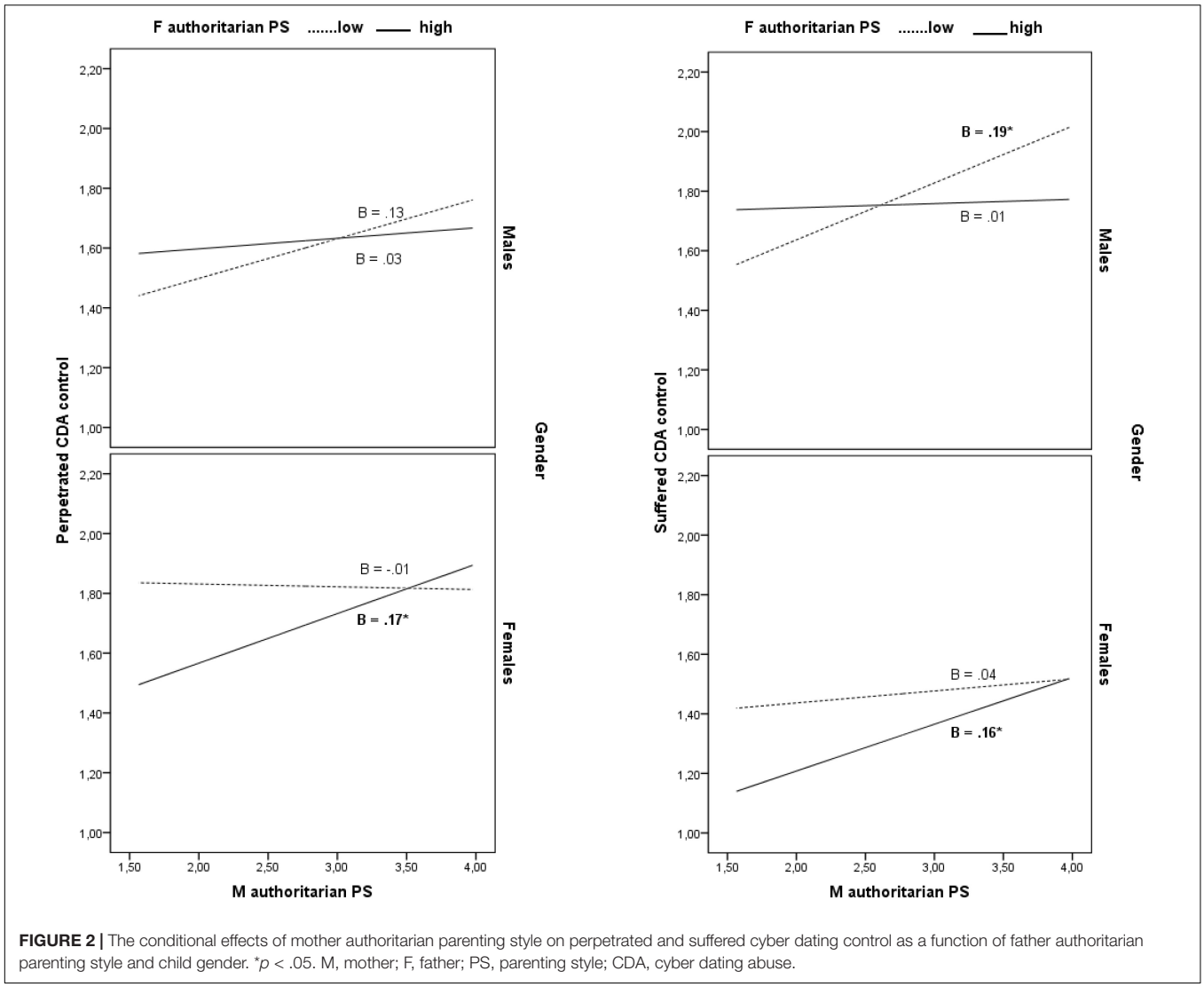


FIGURE 2 | The conditional effects of mother authoritarian parenting style on perpetrated and suffered cyber dating control as a function of father authoritarian parenting style and child gender. *$p < .05$. M, mother; F, father; PS, parenting style; CDA, cyber dating abuse.

tend to be more impulsive, lacking self-regulation and self-control (Patock-Peckham et al., 2001; Piotrowski et al., 2013). These features might expose them to a higher risk of acting aggressively toward their partner not only offline (Pinquart, 2017) but also online, and of overestimating and poorly bearing their partner's control.

Regarding authoritative parenting, contrary to the literature on offline externalizing and abusive behaviors (Pinquart, 2017), but consistent with a growing line of research that questions the protective role of the authoritative style in relation to cyber aggression (Muñiz-Rivas et al., 2019; Zhang et al., 2021), we found that this parenting style was not uniquely related to children's CDA perpetration and victimization. Possibly, other variables which were not considered in this study, such as parent-child communication about affective relations and risks and opportunities of new technologies, might moderate the relationship between authoritative parenting and CDA.

Overall, mothers' authoritarian and permissive parenting practices related more strongly to their children's involvement in cyber abusive relationships than fathers' parenting practices. This finding supports our hypothesis (H2) and Muñiz-Rivas et al.'s (2019) results and can be explained by the primary role mothers are expected to play in children rearing, especially in the areas of effective relationships. Indeed, consistent with the dominant gendered expectations and the ideology of "intensive mothering" (Hays, 1996), mothers are, willingly or not, still the primary caregiver in the family (Raley et al., 2012; Carlson et al., 2016) and feel to be the main responsible for their children development and outcomes. Fathers generally have less responsibility for their adolescent children's discipline, daily care, and recreational activities and are also less involved in their children's peer relations (Updegraff et al., 2001; Phares et al., 2009). This evidence calls for a more egalitarian upbringing.

The unique relations of mothers' and fathers' parenting styles with their children perpetrated and suffered CDA were not moderated by children gender, thereby disconfirming our hypothesis (H3) and suggesting that other factors may explain gender differences in CDA, such as hegemonic masculinity and

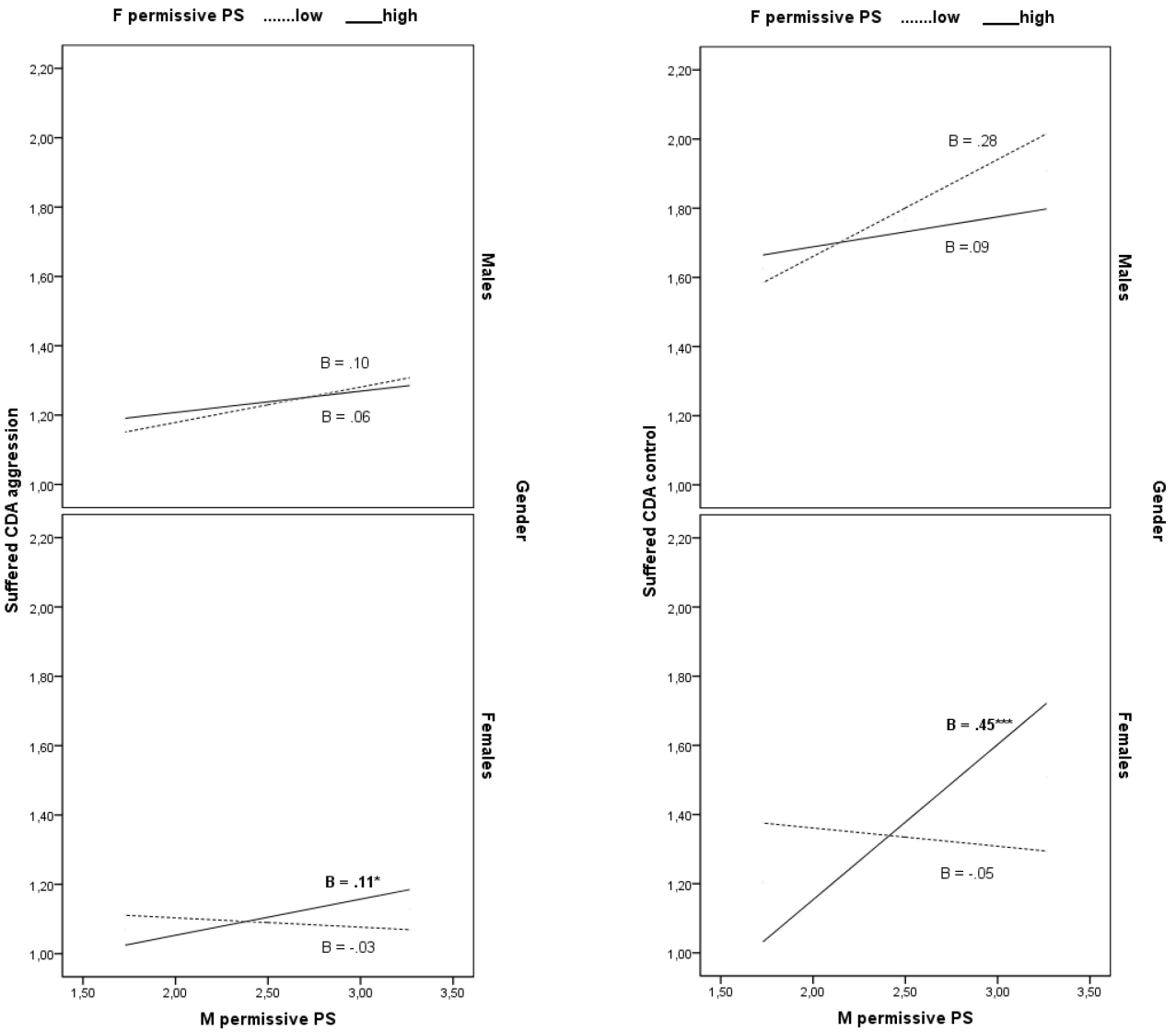


FIGURE 3 | The conditional effects of mother permissive parenting style on suffered cyber dating aggression and control as a function of father permissive parenting style and child gender. *$p < .05$, ***$p < .001$. M, mother; F, father; PS, parenting style; CDA, cyber dating abuse.

sexual aggression myths (March et al., 2021). This result is consistent with a recent meta-analysis that found no moderating effect of gender on the relationship between parenting styles and children's offline externalizing problems (Pinquart, 2017).

Even though not uniquely associated to their children's CDA, fathers' parenting styles do interact with mothers' parenting styles in relating to their daughters' CDA. Specifically, mothers' authoritarian style positively related to their daughters perpetrated and suffered cyber control only if fathers were authoritarian; similarly, mothers' permissive style was positively related to their daughters suffered cyber aggression and control only if fathers were permissive. Consistent with previous evidence (McKinney and Renk, 2008), these findings suggest that congruence in parenting is not necessarily related to beneficial outcomes: when fathers and mothers consistently adopt dysfunctional parenting strategies, their daughters, who usually internalize parents' standards, values, and viewpoints more than sons do (Zentner and Renaud, 2007), might be exposed to a higher risk of perpetrating and suffering CDA.

Moreover, the mothers' authoritarian style was positively related to their daughters perpetrated cyber aggression and to their sons suffered cyber control only if fathers were, respectively, permissive and poorly authoritarian. Therefore, in line with previous research (Ruiz-Hernández et al., 2019), parental inconsistency in parenting styles seems to have detrimental implications for the involvement of children in cyber abusive relationships, especially when it combines two dysfunctional parenting styles.

When interpreting these results, several limitations of the study and avenues for future research should be considered. First,

the small sizes of effects call for larger and more heterogeneous samples to reach more definitive and generalizable conclusion. Second, the cross-sectional design does not provide information on the direction of effects, to explore the which will be important to collect longitudinal data. Third, the children's retrospective perceptions of parenting practices may be different from those actually implemented, therefore the use of observational measures or multi-informant reports that assess parenting practices when they display should be preferred in the future. Finally, given that the different families to which daughters and sons belong may be a confounder of the gender differences that emerged, data provided by male and female siblings from the same family should be collected to reach a better understanding of these differences.

Notwithstanding these limitations, this study made significant contributions to the literature on the role of gendered-differentiated family socialization in the development of cyber abusive romantic relationships in young adulthood. In particular, it shows that specific maternal and paternal parenting styles have not only unique but also complex joint relations with cyber dating aggression and control perpetrated and suffered by their children and that these relations significantly differ across sons and daughters. These findings have also interesting practical implications for educational programs aimed at improving parenting style (for a review see Ryan et al., 2017). Specifically, they suggest that such programs might be more effective when they not only involve both parents but also intervene on each parent's style according to the other parent's style and to the child's sex. Our results might also help parents to become more aware of the wide-ranging impact of their parenting practices on children's offline and online behaviors, and more motivated to get involved in parenting interventions when offered to them.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available on request to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FP and DG designed the study. MP, LC, and DG collected the data, and FP analyzed them. All co-authors participated in the discussion of the results, drafted the manuscript, and approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.818607/full#supplementary-material

REFERENCES

Aguinis, H., Beaty, J. C., Boik, R. J., and Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: a 30-year review. *J. Appl. Psychol.* 90, 94–107. doi: 10.1037/0021-9010.90.1.94

Associazione Italiana di Psicologia [AIP] (2015). *Code of Ethics of the Italian Psychological Association*. College Park, MD: AIP.

Bandura, A. (1977). *Social Learning Theory*. New York: General Learning Press.

Baumrind, D. (1971). Current patterns of parental authority. *Dev. Psychol.* 4, 1–103. doi: 10.1037/h0030372

Baumrind, D. (1991). The influence of parenting style on adolescent competence and substance use. *J. Early Adolesc.* 11, 56–95. doi: 10.1177/0272431691111004

Bennet, D. C., Guran, E. L., Ramos, M. C., and Margolin, G. (2011). College students' electronic victimization in friendships and dating relationships: anticipated distress and associations with risky behaviors. *Violence Vict.* 26, 410–429. doi: 10.1891/0886-6708.26.4.410

Berkien, M., Louwerse, A., Verhulst, F., and van der Ende, J. (2012). Children's perceptions of dissimilarity in parenting styles are associated with internalizing and externalizing behavior. *Eur. Child Adolesc. Psychiatry* 21, 79–85. doi: 10.1007/s00787-011-0234-9

Borrajo, E., Gamez-Gaudix, M., and Calvete, E. (2015). Cyber dating abuse: prevalence, context, and relationship with offline dating aggression. *Psychol. Rep.* 116, 565–585. doi: 10.2466/21.16.PR0.116k22w4

Bowlby, J. (1969). *Attachment and Loss: Attachment*, Vol. 1. New York: Basic Books.

Buck, N. M. L., Leenaars, E. P. E. M., Emmelkamp, P. M. G., and van Marle, H. J. C. (2012). Explaining the relationship between insecure attachment and partner abuse: the role of personality characteristics. *J. Interpers. Violence* 27, 3149–3170. doi: 10.1177/0886260512441258

Buckingham, D., and Willett, R. (2006). *Digital Generations: Children, Young People, and the New Media (1sted.)*. Milton Park: Routledge.

Carlson, D., Hanson, S., and Fitzroy, A. (2016). The division of child care, sexual intimacy, and relationship quality in couples. *Gend. Soc.* 30, 442–466. doi: 10.1177/0891243215626709

Celsi, L., Paleari, F. G., and Fincham, F. D. (2021). Adverse childhood experiences and early maladaptive schemas as predictors of cyber dating abuse: an actor-partner interdependence mediation model approach. *Front. Psychol.* 12:623646. doi: 10.3389/fpsyg.2021.623646

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum.

Cuccì, G., O'Leary, K. D., Olivari, M. G., Bonanomi, A., and Confalonieri, E. (2019). Adolescent dating violence perpetration, emotion dysregulation, and parenting styles. *J. Fam. Psychol.* 33, 12–22. doi: 10.1037/fam0000464

Curry, T. R., and Zavala, E. (2020). A multi-theoretical perspective on cyber dating abuse victimization and perpetration within intimate relationships: a test of general strain, social learning, and self-control theories. *Vict. Offenders* 15, 499–519. doi: 10.1080/15564886.2020.1734996

Darling, N., and Steinberg, L. (1993). Parenting style as context: an integrative model. *Psychol. Bull.* 113, 487–496. doi: 10.1037/0033-2909.113.3.487

Drauker, C. B., and Martsolf, D. S. (2010). The role of electronic communication technology in adolescent dating violence. *J. Child Adolesc. Psychiatr. Nurs.* 3, 133–142. doi: 10.1111/j.1744-6171.2010.00235.x

Elsaesser, C., Russell, B., Ohannessian, C. M., and Patton, D. (2017). Parenting in a digital age: a review of parents' role in preventing adolescent cyberbullying. *Aggress. Violent Behav.* 35, 62–72. doi: 10.1016/j.avb.2017.06.004

Flach, R. M. D., and Deslandes, S. F. (2017). Cyber dating abuse in affective and sexual relationships: a literature review. *Cad. Saude. Publica* 33:e00138516. doi: 10.15090/0102-311X00138516

Guerrero, L. K. (1998). Attachment style differences in the experience and expression of romantic jealousy. *Pers. Relationships* 5, 273–291. doi: 10.1037//0022-3514.72.3.627

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.

Hays, S. (1996). *The Cultural Contradictions of Motherhood*. New Haven, CT: Yale University Press.

He, D., Shen, X., Yang, H., and Fan, C. Y. (2017). The e?ect of parenting style on adolescents' cyber bullying: the mediating role of empathy. *China J. Health Psychol.* 25, 401–406. doi: 10.13342/j.cnki.cjhp.2017.03.023

Junco, R., and Mastrodicasa, J. (2007). *Connecting to the Net.Generation: What Higher Education Professionals Need to Know About Today's Students*. Washington, D.C: NASPA.

Kenny, D. A. (2018). *Moderation*. Available Online at: http://davidakenny.net/cm/moderation.htm (accessed September 15).

Luk, J. W., Patock-Peckham, J. A., Medina, M., Terrell, N., Belton, D., and King, K. M. (2016). Bullying perpetration and victimization as externalizing and internalizing pathways: a retrospective study linking parenting styles and self-esteem to depression, alcohol use, and alcohol-related problems. *Subst. Use Misuse* 51, 113–125. doi: 10.3109/10826084.2015.1090453

McKinney, C., and Renk, K. (2008). Differential parenting between mothers and fathers: implications for late adolescents. *J. Fam. Issues* 29, 806–827. doi: 10.1177/0192513X07311222

March, E., Grieve, R., Clancy, E., Klettke, B., van Dick, R., and Hernandez Bark, A. S. (2021). The role of individual differences in cyber dating abuse perpetration. *Cyberpsychol. Behav. Soc. Netw.* 24, 457–463. doi: 10.1089/cyber.2020.0687

Martínez, I., Garcia, F., Fuentes, M., Veiga, F., Garcia, O., Rodrigues, Y., et al. (2019). Researching parental socialization styles across three cultural contexts: scale ESPA29 Bi-dimensional validity in Spain, Portugal, and Brazil. *Int. J. of Environ. Res. Public Health* 16:197. doi: 10.3390/ijerph16020197

Martínez-Ferrer, B., León-Moreno, C., Musitu-Ferrer, D., Romero-Abrio, A., Callejas-Jerónimo, J. E., and Musitu-Ochoa, G. (2019). Parental socialization, school adjustment and cyber-aggression among adolescents. *Int. J. Environ. Res. Public Health* 16:4005. doi: 10.3390/ijerph16204005

Michiels, D., Grietens, H., Onghena, P., and Kuppens, S. (2008). Parent–child interactions and relational aggression in peer relationships. *Dev. Rev.* 28, 522–540. doi: 10.1016/j.dr.2008.08.002

Mikulincer, M., and Shaver, P. R. (2010). *Attachment in Adulthood: Structure, Dynamics, and Change*. New York, NY: Guilford.

Moreno Méndez, J. H., Espada Sánchez, J. P., and Gómez Becerra, M. I. (2020). Role of parenting styles in internalizing, externalizing, and adjustment problems in children. *Salud Ment.* 43, 73–84. doi: 10.17711/sm.0185-3325.2020.011

Moreno-Ruiz, D., Martínez-Ferrer, B., and Gracìa-Bacete, F. (2019). Parenting styles, cyberaggression, and cybervictimization among adolescents. *Comput. Hum. Behav.* 93, 252–259. doi: 10.1016/j.chb.2018.12.031

Mumford, E. A., Liu, W., and Taylor, B. G. (2016). Parenting profiles and adolescent dating relationship abuse: attitudes and experiences. *J. Youth Adolesc.* 45, 959–972. doi: 10.1007/s10964-016-0448-8

Muñiz-Rivas, M., Vera, M., and Povedano-Díaz, A. (2019). Parental style, dating violence and gender. *Int. J. Environ. Res. Public Health* 16:2722. doi: 10.3390/ijerph16152722

Olivari, M. G., Cuccì, G., and Confalonieri, E. (2017). Stile genitoriale autoritario, giustificazione della violenza e messa in atto di comportamenti di dating aggression. *Maltrattam. Abus Infanz.* 19, 13–29. doi: 10.3280/mal2017-003002

Patock-Peckham, J. A., Cheong, J., Balhorn, M. E., and Nagoshi, C. T. (2001). A social learning perspective: a model of parenting styles, self regulation, perceived drinking control, and alcohol use and problems. *Alcohol. Clin. Exp. Res.* 25, 128–1292. doi: 10.1111/j.1530-0277.2001.tb02349.x

Phares, V., Fields, S., and Kamboukos, D. (2009). Fathers' and mothers' involvement with their adolescents. *J. Child Fam. Stud.* 18, 1–9. doi: 10.1007/s10826-008-9200-7

Pinquart, M. (2017). Associations of parenting dimensions and styles with externalizing problems of children and adolescents: an updated meta-analysis. *Dev. Psychol.* 53, 873–932. doi: 10.1037/dev0000295

Piotrowski, J. T., Lapierre, M. A., and Linebarger, D. L. (2013). Investigating correlates of self-regulation in early childhood with a representative sample of english-speaking american families. *J. Child Fam. Stud.* 22, 423–436. doi: 10.1007/s10826-012-9595-z

Raley, S., Bianchi, S. M., and Wang, W. (2012). When do fathers care? mothers' economic contribution and fathers' involvement in childcare. *Am. J. Sociol.* 117, 1422–1459. doi: 10.1086/663354

Ramos, M. C., Miller, K. F., Moss, I. K., and Margolin, G. (2017). Perspectivetaking and empathy mitigate family-of-origin risk for electronic aggression perpetration toward dating partners: a brief report. *J. Interpers. Violence* 18, 1155–1164. doi: 10.1177/0886260517747605

Reed, L. A., Tolman, R. M., and Ward, L. M. (2017). Gender matters: experiences and consequences of digital dating abuse victimization in adolescent dating relationships. *J. Adolesc.* 53, 79–89. doi: 10.1016/j.adolescence.2017.05.015

Reid, J. B., Patterson, G. R., and Snyder, J. (2002). *Antisocial Behavior in Children and Adolescents: a Developmental Analysis and Model for Intervention*. Washington, DC: American Psychological Association. doi: 10.1037/10468-000

Robinson, C. C., Mandleco, B., Olsen, S. F., and Hart, C. H. (2001). "The Parenting Styles and Dimensions Questionnaire (PSDQ)," in *Handbook of Family Measurement Techniques, Instruments and Index*, eds B. F. Permultter, J. Touliatos, and G. W. Holden (Thousand Oaks, CA: Sage), 319–321.

Ruiz-Hernández, J. A., Moral-Zafra, E., Llor-Esteban, B., and Jiménez-Barbero, J. A. (2019). Influence of parental styles and other psychosocial variables on the development of externalizing behaviors in adolescents: a systematic review. *Eur. J. Psychol. Appl. Leg. Context.* 11, 9–21. doi: 10.5093/ejpalc2018a11

Sargent, K. S., Krauss, A., Jouriles, E. N., and McDonald, R. (2016). Cyber victimization, psychological intimate partner violence, and problematic mental health outcomes among first-year college students. *Cyberpsychol. Behav. Soc. Netw.* 19:54550. doi: 10.1089/cyber.2016.0115

Smith-Darden, J. P., Reidy, D. E., and Kernsmith, P. D. (2016). Adolescent stalking and risk of violence. *J. Adolesc.* 52, 191–200. doi: 10.1016/j.adolescence.2016.08.005

Soenens, B., Vansteenkiste, M., and van Petegem, S. (2015). Let us not throw out the baby with the bathwater: applying the principle of universalism without uniformity to autonomy-supportive and controlling parenting. *Child Dev. Perspect.* 9, 44–49.

Suárez-Relinque, C., Del Moral, Arroyo, G., León-Moreno, C., and Callejas Jerónimo, J. E. (2019). Child-to-parent violence: which parenting style is more protective? a study with Spanish adolescents. *Int. J. Environ. Res. Public. Health* 16:1320.

Ryan, R., O'Farrelly, C., and Ramchandani, P. (2017). Parenting and child mental health. *London J. Prim. Care* 9, 86–94.

Tagliabue, S., Olivari, M. G., Bacchini, D., Affuso, G., and Confalonieri, E. (2014). Measuring adolescents' perceptions of parenting style during childhood: psychometric properties of the parenting styles and dimensions questionnaire. *Psicol. Teoria e Pesquisa* 30, 251–258.

Updegraff, K. A., McHale, S. M., Crouter, A. C., and Kupanoff, K. (2001). Parents' involvement in adolescents' peer relationships: a comparison of mothers' and fathers' roles. *J. Marriage Fam.* 63, 655–668. doi: 10.1111/j.1741-3737.2001.00655.x

Van Ouytsel, J., Torres, E., Choi, H. J., Ponnet, K., Walrave, M., and Temple, J. R. (2017). The associations between substance use, sexual behaviors, bullying, deviant behaviors, health, and cyber dating abuse perpetration. *J. Sch. Nurs.* 33, 116–122. doi: 10.1177/1059840516683229

Vale, A., Pereira, F., Gonçalves, M., and Matos, M. (2018). Cyber-aggression in adolescence and internet parenting styles: a study with victims, perpetrators and victim-perpetrators. *Child. Youth Serv. Rev.* 93, 88–89. doi: 10.1016/j.childyouth.2018.06.021

World Medical Association (2013). World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194.

Zentner, M., and Renaud, O. (2007). Origins of adolescents' ideal self: an intergenerational perspective. *J. Pers. Soc. Psychol.* 92, 557–574. doi: 10.1037/0022-3514.92.3.557

Zhang, Y., Chen, C., Teng, Z., and Guo, C. (2021). Parenting style and cyber-aggression in chinese youth: the role of moral disengagement and moral identity. *Front. Psychol.* 12:621878. doi: 10.3389/fpsyg.2021.621878

Zhao, F., and Gao, W. (2012). Reliability and validity of adolescent online aggressive behavior scale. *Chin. Mental Health J.* 26, 439–444.

Zweig, J. M., Dank, M., Yahner, J., and Lachman, P. (2013). The rate of cyber dating abuse among teens and how it relates to other forms of teen dating violence. *J. Youth Adolesc.* 42, 1063–1077. doi: 10.1007/s10964-013-9922-8

Zweig, J. M., Lachman, P., Yahner, J., and Dank, M. (2014). Correlates of cyber dating abuse among teens. *J. Youth Adolesc.* 43, 1306–1321. doi: 10.1007/s10964-013-0047-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fpsyg.2022.794896

Self-Control Capacity Moderates the Effect of Stereotype Threat on Female University Students' Worry During a Math Performance Situation

Alex Bertrams[1], Christoph Lindner[2,3], Francesca Muntoni[2] and Jan Retelsdorf[2]*

[1]Educational Psychology Lab, Institute of Educational Science, University of Bern, Bern, Switzerland, [2]Faculty of Education, University of Hamburg, Hamburg, Germany, [3]Leibniz Institute for Science and Mathematics Education, Kiel, Germany

Stereotype threat is a possible reason for difficulties faced by girls and women in the fields of science, technology, engineering, and mathematics. The threat experienced due to gender can cause elevated worry during performance situations. That is, if the stereotype that women are not as good as men in math becomes salient, this stereotype activation draws women's attention to task-irrelevant worry caused by the fear of conforming to the negative stereotype. Increased worry can reduce cognitive resources, potentially leading to performance decrements. We argue that such worry is more pronounced immediately after an unrelated self-control demand, which is assumed to temporarily decrease people's self-control exertion over their attention and stream of thought (i.e., relatively low self-control capacity). This prediction was examined in an experiment conducted with 102 participating university students enrolled in courses in which math plays a crucial role. After the manipulation of self-control capacity (low vs. high), stereotype threat was induced for the female students, but not the male students. Then, the students were asked to report their thoughts during a math performance situation (i.e., written thought protocols) three times. Multiple-group autoregressive path models revealed that when self-control capacity was relatively low, female compared with male students reported more intense worry in the initial two thought protocols. In contrast, in the relatively high self-control capacity condition, female and male students did not differ significantly in their reported worry at any time. These results expand on previous findings, suggesting that threat effects depend on definable situational self-control conditions.

Keywords: gender, self-control capacity, self-regulation, stereotype threat, test anxiety, worry

OPEN ACCESS

Edited by:
Nahia Idoiaga, University of the Basque Country, Spain

Reviewed by:
Verónica Romero-Ferreiro, Complutense University of Madrid, Spain
Tobias Tempel, University of Trier, Germany

****Correspondence:***
Alex Bertrams
alexander.bertrams@edu.unibe.ch

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *14 October 2021*
Accepted: *09 March 2022*
Published: *07 April 2022*

Citation:
Bertrams A, Lindner C, Muntoni F and Retelsdorf J (2022) Self-Control Capacity Moderates the Effect of Stereotype Threat on Female University Students' Worry During a Math Performance Situation. Front. Psychol. 13:794896. doi: 10.3389/fpsyg.2022.794896

INTRODUCTION

The underrepresentation of girls and women in the science, technology, engineering, and mathematics (STEM) fields continues to be a concern not only for educators and scientists, but also for society. Gender differences in mathematics and science achievement have been reported in the literature over many decades (Hedges and Nowell, 1995; Reilly et al., 2019). Furthermore, boys also report more positive attitudes toward learning mathematics and science

(Else-Quest et al., 2010), and girls report lower self-confidence and greater mathematics anxiety (Beilock et al., 2010). However, researchers are divided on how significant gender differences in mathematics and science are. Some argue that the differences are small but still meaningful (Reilly et al., 2015), while others argue that these differences are very minimal on average or do not exist (Hyde, 2005; Else-Quest et al., 2010). For example, the gender similarities hypothesis, proposed by Hyde (2005), states that males and females are similar on most, but not all, psychological variables, and meta-analytic evidence suggests that girls and boys do not, in fact, perform differently in measures of mathematics achievement (Lindberg et al., 2010). Nevertheless, from an early age, children show rigid gender stereotyping, perceiving mathematics and science as male domains (Plante et al., 2009; Martinot et al., 2012). These negative stereotypes can affect girls' mathematics performance through the mechanism of stereotype threat (Evans et al., 2011; Tomasetto et al., 2011; Passolunghi et al., 2014). Moreover, it has been found that these gender stereotypes persist into adulthood and are a cross-cultural phenomenon (Nosek et al., 2002).

Stereotype threat has frequently been studied as a possible reason for the weaker engagement, motivation, and performance of girls and women in STEM (Pennington et al., 2016; Spencer et al., 2016). The stereotype that females are less able than males in STEM tasks can cause fear of confirming the stereotype, leading to task-irrelevant worry and even performance degradation in STEM test situations (Schmader and Johns, 2003). To the best of our knowledge, no study has yet examined the moderating situational circumstances of the stereotype threat effect on worry. Addressing this lacuna, in this study, we examine whether worry is more pronounced immediately after an unrelated effortful self-control demand which causes relatively low momentary self-control capacity. To do so, we drew on a university student sample enrolled in courses in which math plays a crucial role. The current study allowed us to extend previous research in at least two ways. First, in addition to previous research focusing on the relation between stereotype threat and worry (e.g., Cadinu et al., 2005), we considered the possible moderating function of students' self-control capacity. Second, this study augments the existing research on test anxiety and self-control capacity (e.g., Bertrams et al., 2013) by investigating the effect of another origin of threat (i.e., being stereotyped).

Stereotype Threat and the Role of Worry

A growing body of literature has investigated the relation between stereotypes and performance. The most significant work in this area is the research on *stereotype threat*, which indicates that negative stereotypes hamper the academic performance of stereotyped individuals (Steele and Aronson, 1995; Tempel and Neumann, 2014; Pennington et al., 2016; Spencer et al., 2016). The stereotype threat effect has been characterized as a "psychological predicament in which individuals are inhibited from performing to their potential by the recognition that possible failure could confirm a negative stereotype that applies to their ingroup and, by extension, to themselves" (Schmader, 2002, p. 194). Thus, stereotype threat is viewed as a form of social identity threat.

In their influential paper, Steele and Aronson (1995) noted that the stereotype that African American students have lower academic ability hampered the performance of African American students on academic tests. As African Americans are well aware of the negative stereotypes questioning their intellectual ability, they might fear confirming this stereotype. This fear of stereotype confirmation is considered to occupy the cognitive systems required for optimal performance and therefore leads to low test performance. This study has stimulated numerous subsequent studies investigating the influence of negative stereotypes. For example, when confronted with the negative stereotype about their in-group, women were found to underperform on math tests (e.g., Spencer et al., 1999) and driving tests (Yeung and von Hippel, 2008), older adults were found to underperform on memory tests and cognitive tests (Lamont et al., 2015), and students with lower socio-economic status were found to underperform on intelligence tests (Désert et al., 2009). However, of the many negative stereotypes that have been studied in the context of stereotype threat, the stereotype that women are not as good in mathematics as men is one of the most frequently studied (e.g., Spencer et al., 1999; Schmader, 2002; Tomasetto et al., 2011; Tomasetto, 2019). Research on this topic has shown that women and girls exhibit lower math performance if they are reminded of the negative stereotype about girls and their inferior math ability, but they perform as well as boys if such stereotypes are not made salient before they take a math test.

However, what causes performance declines in stereotype-threatening situations? Stereotype threats do not generally lead to decreased motivation in performance situations. Instead, people experiencing stereotype threat are motivated to disprove negative stereotypes about their social identity (e.g., Nussbaum and Steele, 2007; Vandello et al., 2008) or at least to avoid confirming them (e.g., Brodish and Devine, 2009; Chalabaev et al., 2012). Thus, stereotype threat creates the desire to do well on a given task and disprove the negative stereotypes (Steele and Aronson, 1995). The motivation to disconfirm the stereotype, or to avoid confirming it, represents pressure to succeed; however, high effort cannot always be invested. Instead, higher motivation to do well in stereotype-threatening situations can produce distracting and negative thoughts (Schmader et al., 2008).

According to Schmader et al.'s (2008) integrated process model of stereotype threat, when a negative stereotype becomes relevant to one's performance, it triggers increased physiological arousal, which impairs working memory operations. Furthermore, stereotyped individuals are busy detecting self-relevant information and signs of failure, which is the second process that puts a strain on working memory. The last process is the suppression of negative thoughts and feelings resulting from the first two processes, which further consumes the working memory capacity necessary for successful performance. Taken together, all three processes described above lead to reduced working memory capacity in tasks requiring cognitive resources, which can lead to uncharacteristically poor performance on

a test. In fact, Schmader and Johns (2003) found that priming negative stereotypes reduced women's memory capacity. Furthermore, they determined that a reduction in working memory capacity mediated the effect of stereotype threat on women's math performance. These results are in line with the idea that individuals experiencing stereotype threats have negative thoughts, which reduce their working memory capacity. Numerous studies conducted subsequently demonstrated negative cognition in stereotype threat situations, such as negative expectancies and thoughts (Cadinu et al., 2005) and task-related worry (Beilock et al., 2007; Gerstenberg et al., 2012). To concentrate on a test, these worry thoughts need to be ignored, which consumes cognitive resources.

Applied to the field of women in math, this means that if the stereotype that women are not as good as men in math becomes salient in a test situation, this stereotype activation draws women's attention to task-irrelevant worry caused by the fear of conforming to the negative stereotype. Increased worry then reduces cognitive resources, thus degrading performance. In fact, in previous studies, women who were told that gender differences in math exist (i.e., stereotype threat) not only performed worse, but also reported having more negative thoughts about math compared with a control group (Cadinu et al., 2005; Beilock et al., 2007; Gerstenberg et al., 2012). In particular, Cadinu et al. (2005) asked a group of female university students to complete a difficult math test under stereotype threat or in a no-threat (control) condition. During the task, the women were asked to list any thoughts that came to their mind immediately before solving each of the seven difficult math problems. The authors predicted, *inter alia*, that individuals under stereotype threat would report more worry thoughts and show a decrease in performance compared with those under the control condition, and that worry would mediate the negative effects of stereotype threat on performance. As predicted, Cadinu et al.'s study found that women under stereotype threat reported more worry thoughts related to the test and showed a sharp decrease in performance compared with those in the no-threat condition. More importantly, performance degradation was mediated by an increase in worry. Therefore, they concluded that negative performance-related thoughts can consume working memory capacities to impede performance.

Similarly, Beilock et al. (2007) found that stereotype threat resulted in a greater proportion of task-related worry. Furthermore, this relation was attributed to the consumption of working memory resources. In another study conducted by Gerstenberg et al. (2012), female university students were asked about their current thoughts before the math test, but after the stereotype manipulation. They showed the highest level of worry thoughts (e.g., "I ask myself whether my performance will be good enough.") when the stereotype threat was activated. Moreover, as has been shown by other researchers (e.g., Cadinu et al., 2005; Beilock et al., 2007), performance-inhibiting worry mediated the stereotype threat effect.

In summary, as Schmader et al.'s (2008) integrated process model suggests, if attention shifts away from the task, it is because people are having (negative) thoughts about their performance. Thus, in stereotype-threatening situations, distracting worry thoughts reduce the cognitive resources required to successfully elaborate task-relevant information.

Self-Control Capacity in Test Situations

Self-control is defined as the mental capacity that enables people to override, inhibit, or modify their impulses, emotions, thoughts, and behaviors and to bring them in line with standards and personally endorsed overarching goals (e.g., Baumeister et al., 2007). Studies in different fields of psychological research have provided evidence that dealing with initial self-control demands briefly undermines an individual's cognitive capacities that are required for subsequent working memory operations, regulating thoughts and emotions, and focusing attention in a goal-directed manner (Baumeister and Vohs, 2016). There is an ongoing debate regarding which mechanism underlies the effect of self-control-dependent performance decrements (e.g., Kurzban et al., 2013; Bertrams, 2020); in this regard, some authors have associated the detrimental effects of initial self-control on subsequent operations with mental fatigue or exhaustion (e.g., Job et al., 2010; Bertrams, 2020). It is also debated whether such an exhaustion effect of self-control exists at all (e.g., Englert and Bertrams, 2021). However, there is reasonable theory and empirical evidence for the existence of a varying self-control capacity, and many researchers have agreed that self-control cannot always be maintained in cognitively demanding situations (Baumeister and Vohs, 2016; Garrison et al., 2019; Bertrams, 2020; Dang et al., 2020).

In achievement tests, the exertion of self-control is required for cognitive processing as well as for focusing on the items' content over longer periods while inhibiting distractions (e.g., negative thoughts). More precisely, executive functions [i.e., updating of working memory, inhibiting impulses, and shifting between mental sets; (Miyake et al., 2000)] seem to be key ingredients for successful self-control in achievement situations (Hofmann et al., 2012). Students with low compared with high levels of self-control capacity perform worse in working memory (Schmeichel, 2007), logical reasoning, and mental arithmetic tasks (Schmeichel et al., 2003).

Englert and Bertrams (2017) showed that eighth graders' knowledge retrieval was undermined when their self-control capacity was briefly depleted in a previous unrelated self-control demanding task. This result indicates the relation between low self-control capacity and low working memory capacity, two ingredients that are required for working focused on item content, especially in science and math tests. In line with this assumption, Lindner et al. (2019) found that students with low working memory capacities showed an early onset of rapid-guessing behavior (i.e., unrealistic fast responses to test items) over the course of a science test, indicating a reduction in students' test-taking efforts. In another study (Lindner et al., 2017), students with lower self-control capacity showed stronger progressive performance declines over the course of a computer-based mathematical problem-solving test compared with individuals with higher levels of self-control capacity. It can be assumed that the performance decrements were due to an increasing number of distracting thoughts, which could not be regulated effectively during the testing procedure.

As mentioned above, self-control capacity is also required for regulating negative emotions and thoughts. Consistent with this notion, Lindner and Retelsdorf (2020) found that students who perceived themselves as having lower levels of actual self-control capacity subsequently showed lower scores in an English as a foreign language test and reported more cognitive interruptions due to distracting thoughts. More directly related to threat and related worry during test situations, Bertrams et al. (2013) showed that test-anxious students who are more susceptible to experiencing threat were distracted more frequently by anxiety-related worry thoughts and, therefore, performed worse in an arithmetic test. However, the relational pattern among test anxiety, worry, and performance was found only in individuals whose self-control capacity had initially been experimentally impaired, not in individuals whose self-control capacity was intact.

Present Research

Based on previous research, we assumed that the fear of stereotype confirmation can occupy the capacity of the cognitive system, and stereotype-threatening situations might trigger distracting and negative thoughts (Schmader et al., 2008). However, stereotype threat during evaluative situations should be more strongly related to distracting worry thoughts when the self-control capacity is momentarily lower compared with being intact. This pattern resembles Bertrams et al.'s (2013) findings on test anxiety, self-control capacity, and worry, as test anxiety is considered to be associated with the experience of threat (Spielberger and Vagg, 1995). However, unlike the existing research, we examined the effect of another cause of threat (i.e., being stereotyped) than trait test anxiety on worry with regard to self-control capacity.

More precisely, we examined the moderating influence of self-control capacity on gender stereotype threat in predicting the development of worry thoughts prior to and during an evaluative math test. For this reason, we invited female and male students, all of whom studied math at a German university, to the lab. The self-control capacity was impaired in the experimental condition and left intact in the control condition. In addition, we induced stereotype threat for all female but not male participants through standardized test instruction. Then, the participants reported their thoughts three times (after the stereotypical test instruction and after each half of a brief alleged test of mental arithmetic abilities). We assumed that being female is associated with higher worry in the present performance situation (Cadinu et al., 2005). However, based on previous research (Bertrams et al., 2013), we predicted that the relation between gender and worry would be more pronounced when self-control capacity is low compared with high.

MATERIALS AND METHODS

Participants

Overall, 104 undergraduates enrolled in studies related to math (business mathematics or teaching training for math in schools) at a German university participated. None of them correctly guessed the true purpose of this study, but two indicated that they did not speak German fluently. As all materials in this study were presented in German, we decided prior to the analyses not to include their data. Thus, the final sample comprised 102 students (58% female; $M_{age} = 21.70$, $SD_{age} = 2.29$). The participants were randomly assigned to either the low ($n = 50$; $n = 29$ female, 21 male) or the high ($n = 52$; $n = 30$ female, 22 male) self-control capacity groups. The sample size decision was based on Simmons et al.'s (2011) recommendation to collect at least 20 participants per group, a recommendation that was a common guideline at the time the present data were collected (i.e., 2014).

Negative stereotypes about math abilities might not affect individuals who are not skilled or to whom math is essentially unimportant (Steele, 1997). However, the participants' answers on their math skills ($M = 6.13$, $SD = 1.33$) and personal importance to be good at math ($M = 6.51$, $SD = 1.88$) indicated sufficiently high math-related skills and relevance, as they were above the midpoint of nine-point scales (Beilock et al., 2007). Moreover, the mean final grade from the secondary school (German Abitur; $M = 1.82$, $SD = 0.54$) indicated a high cognitive ability in this sample, given that it strongly diverged from the average Abitur grade in Germany in each of the 15 years from 2006 to 2020 (means interval = [2.37, 2.52], $ps < 0.001$, $ds > 1.02$, one-sample t-tests; note that lower numbers indicate higher performance in the German grading system; the German average Abitur grades were retrieved from Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2022, January 7).

Procedure

Participation lasted 25–30 min and occurred in a university laboratory room. After giving informed consent, the participants completed a "questionnaire about mathematics," which included the measure of math-related trait test anxiety, one item on self-rated math ability, and another item on the personal relevance of math. Next, the manipulation of self-control capacity was performed. After that, the participants answered the manipulation check, a measure of self-competence, and a mood scale. This was followed by instructions for the math test and an explanation of the thought protocols. The explanation of the thought protocols also included the induction of stereotype threat for the female participants. Subsequently, the participants were asked to fill in the first thought protocol. After that, they worked on a brief math test that was interrupted by a second thought protocol. After the second part of the math test, the participants completed a third thought protocol. Then, the participants answered a questionnaire on their demographic data. The questionnaire also included items on their motivation to perform well during the math test as well as questions on the school-leaving grade, German-language ability, and hypothesis suspicion. Finally, the participants were thanked, debriefed, and either received course credit or €4 in exchange for their participation.

Materials

Manipulation of Self-Control Capacity

We applied a manipulation task from previous research (e.g., Bertrams et al., 2010; Dummel and Rummel, 2016; Wiesner and Lindner, 2017). All participants were asked to transcribe a historical text about a German city. The text was free from threat-related content (e.g., battles, war, and fear). While the high self-control capacity group transcribed the text as it was (i.e., without further instructions), the low self-control group was instructed to always omit the frequent letters "e" and "n." Thus, only the participants in the latter group had to volitionally override their elaborated writing habits and reduced their momentary self-control capacity by this self-control exertion. An English translation of this task is provided in the supplemental material of Bertrams et al. (2015). The experimenter stopped the participants after 6 min and asked them to put the sheet in a prepared concealing desk tray to remain blind to the experimental condition.

Induction of Stereotype Threat

All participants received a sheet explaining how to work on the thought protocols. The stereotype threat for the female participants was integrated into this instruction. It was claimed that the present research was about the cognitive processes that might explain why males have ostensibly been found to consistently perform better in math than females in the educational field, as well as in standardized lab tests. Thus, the instruction highlighted gender differences to the disadvantage of females. Such procedures have been extensively used to induce stereotype threat (e.g., Cadinu et al., 2005; Beilock et al., 2007).

Math Performance Situation

The participants received a single sheet presenting 20 arithmetic tasks in a row on the left. Each task comprised an initial subtraction, followed by a division. Examples are "(43–27): 8 =," "(41–23): 4 =," and "(43–27): 7 =." Next to each task were 18 columns for indicating the solution by checking one box: "1," "1 with remainder," "2," "2 with remainder," … "9 with remainder." We used this unusual arithmetic performance task, as this study did not focus on performance but on worry thoughts. Therefore, we made an effort to standardize the performance situation as much as possible (e.g., by making it unlikely that participants would use and focus on notes to facilitate mental calculation). The instruction for the performance task was to work as quickly as possible, as the working time was limited to 90 s, as well as to avoid wrong solutions. There were two performance blocks (i.e., two single sheets with 20 tasks each) located among the three thought protocols. For each block, the participants were stopped by the experimenter after 90 s.

Worry During the Math Test (Thought Protocols)

The participants were asked to write down on a lined sheet anything they could remember that went through their mind during the last half minute (first thought protocol)/while working on the block of math tasks they had just worked on (second and third thought protocols). Two independent judges rated the number of test-related worry thoughts mentioned in each thought protocol (Cadinu et al., 2005). The inter-rater reliabilities for the number of worry thoughts reported in the first, second, and third thought protocols were satisfying: *ICC*s = 0.78, 0.90, and 0.87, respectively, all *p*s < 0.001. The same was true for inter-rater reliabilities regarding the provided overall number of thoughts: *ICC*s = 0.96, 0.95, and 0.94, respectively, all *p*s < 0.001. Therefore, we averaged the two judges' counts for each measurement time. Conforming with previous research (e.g., Beilock et al., 2007), we adjusted each individual's number of worry thoughts to the individual overall number of thoughts separately for each measurement time. The resulting proportions were the three dependent variables of the extent of worry at the three measurement times.

We also asked the participants to write next to each thought the percentage of the overall time they spent on the respective content. This explorative measure, to the best of our knowledge, has not been applied in the past. As we received estimates that exceeded 100%, we had doubts about the usefulness of this measurement and decided not to analyze the respective responses further.

Self-Reports

We applied a brief version of the Test Anxiety Inventory-German (Wacker et al., 2008). With nine items, the susceptibility to experience of worry and emotionality during math test situations was assessed (e.g., "I am thinking about the consequences of possible failure," and "My heart is pounding."). The answers were given on a four-point Likert scale ranging from *almost never* (1) to *almost always* (4). McDonald's omega was found to be 0.86.

The manipulation check comprised three items that have been previously used (Bertrams and Englert, 2014): "How effortful did you find the transcription task?," "How difficult did you find it to execute the transcription task?," and "How much did you suppress your usual writing habits during the transcription task?" McDonald's omega across these three items was 0.58. In addition, the participants indicated on one item their self-perceived success regarding the transcription task (Bertrams and Englert, 2014): "How much did you succeed in performing the transcription task?" This measure was used to estimate whether the task could unintentionally influence perceptions of self-competence. These four items were completed on a seven-point Likert scale ranging from *not at all* (1) to *very much* (7).

The participants also completed the 10-item Positive and Negative Affect Schedule (Mackinnon et al., 1999). Five items were momentary positive affect (e.g., "inspired"), and the other five items were momentary negative affect (e.g., "afraid") measured on a scale ranging from *not at all* (1) to *extremely* (5). In the present study, McDonald's omega was 0.74 for positive affect and 0.64 for negative affect.

With two items obtained from Beilock et al. (2007), self-perceived math skills ("I am good at math.") and the importance of math ("It is important to me that I am good at math.")

were measured at the beginning of the experimental procedure. Answers were given on a scale ranging from *does not apply at all* (1) to *does completely apply* (9). Moreover, to estimate overall academic skills, we asked for the average final grade from secondary school (German Abitur), with 1.0 as the best possible grade in Germany. Math achievement is an integral part of the German final secondary school grade.

For each of the two performance blocks, the participants indicated their motivation behind the respective performance block on two items ("I made an effort to solve as many math problems as possible," and "I have made an effort to avoid incorrect answers."). The participants responded to the items on a seven-point Likert scale ranging from *does not apply at all* (1) to *does completely apply* (7).

RESULTS

Preliminary Analyses

With a series of two-way ANOVAs, we performed several comparisons among the four groups, resulting from crossing gender (female vs. male) with the self-control capacity condition (low vs. high). In the present specific sample of math students, math-related trait test anxiety did not vary in dependence on gender, self-control capacity condition, or their interaction ($ps > 0.12$). Therefore, we did not incorporate trait test anxiety into the model structures we empirically tested. Furthermore, there were no differences with respect to gender, self-control capacity condition, or their interaction in the self-reported motivation during the performance blocks ($ps > 0.48$).

As revealed by another two-way ANOVA with the factors gender and self-control capacity condition, the manipulation check indicated that the participants in the low compared with the high self-control capacity condition exerted more self-control during the transcription task ($M = 4.03$, SD = 1.11 vs. $M = 2.85$, SD = 0.99, $F[1, 98] = 29.56$, $p < 0.001$, $\eta^2_{part} = 0.23$). There were no group differences in the manipulation check as a function of gender or the interaction between gender and the self-control capacity condition ($ps > 0.28$). As expected, there was no difference in positive or negative mood immediately after the self-control capacity manipulation in dependence on gender, self-control capacity condition, or their interaction ($ps > 0.18$; two-way ANOVA). These results suggest that the self-control capacity manipulation was successful, while it did not unintentionally affect mood.

Diverging from previous research (e.g., Bertrams et al., 2013), a two-way ANOVA yielded an unexpected interaction between gender and the self-control capacity condition with respect to the item responses on self-perceived competence during the transcription task ($F[1, 98] = 8.00$, $p = 0.006$, $\eta^2_{part} = 0.08$). There was also a main effect of the self-control capacity group ($F[1, 98] = 6.25$, $p = 0.01$, $\eta^2_{part} = 0.06$), but no main effect of gender ($p = 0.20$). To interpret the significant interaction, we conducted independent samples *t*-tests. For female university students, self-perceived competence was lower in the low compared with the high self-control capacity condition ($M = 4.28$, SD = 1.07 vs. $M = 5.47$, SD = 1.04, $t[57] = -4.34$, $p < 0.001$, $d = -1.13$). This was not true for male university students ($M = 4.62$, SD = 1.02 vs. $M = 4.55$, SD = 1.34, $t[41] = 0.20$, $p = 0.84$, $d = 0.06$). Moreover, female compared with male university students reported higher perceived competence in the high self-control capacity condition ($M = 5.47$, SD = 1.04 vs. $M = 4.55$, SD = 1.34, $t[50] = 2.80$, $p = 0.007$, $d = 0.79$), but not in the low self-control capacity condition ($M = 4.28$, SD = 1.07 vs. $M = 4.62$, SD = 1.02, $t[48] = -1.14$, $p = 0.26$, $d = -0.33$). However, perceived competence was neither for female nor for male university students within any of the self-control capacity groups correlated with worry at any time of measurement ($ps > 0.15$). Moreover, including perceived competence as covariate did not change the findings obtained for the more parsimonious model without this covariate (see "Main Analyses"). This suggests that any relation between gender and worry was not attributable to the possibility that the transcription task caused lasting group differences in perceived competence.

Main Analyses

To test our hypotheses regarding the prediction of worry by gender in different experimental conditions, we relied on multiple-group autoregressive path models. Therefore, we estimated the model depicted in **Figure 1**. This approach allowed us to directly test the moderation of the path coefficients from gender to the three worry variables at each time point by the experimental condition (low vs. high self-control capacity). Therefore, we applied the Wald test for each standardized path coefficient to determine whether it differed significantly between the two experimental conditions, indicating a moderation of the path by the participants' state of self-control capacity. As our model comprised manifest variables and all possible relations between the variables were allowed, our model was saturated with $df = 0$, and no model fit statistics could be calculated. The model was estimated using M*plus* 8.4 (Muthén and Muthén, 2017) and applying a robust maximum likelihood estimator.

The means and standard deviations for the worry measures are presented in **Table 1**. As shown in **Figure 1**, in the high self-control capacity condition, there were no significant relations between gender and worry at any measurement time. In contrast, in the low self-control capacity condition, gender significantly predicted worry at the first and second measurement times (i.e., female university students reported more worry than male university students), but not at the third time. The path from gender to worry at the second measurement time (i.e., immediately after the first 90-s experience with the performance situation) was significantly different between the low and high self-control capacity conditions (**Table 2**).

Auxiliary Analyses

In the high self-control capacity condition, the worry measures at the different measurement times were significantly related to each other, except for the worry at the second and third measurement times (**Figure 1**). This was different in the low self-control capacity condition, where worry at the first measurement time was not significantly related to worry measures at the second and third times. However, the moderation of

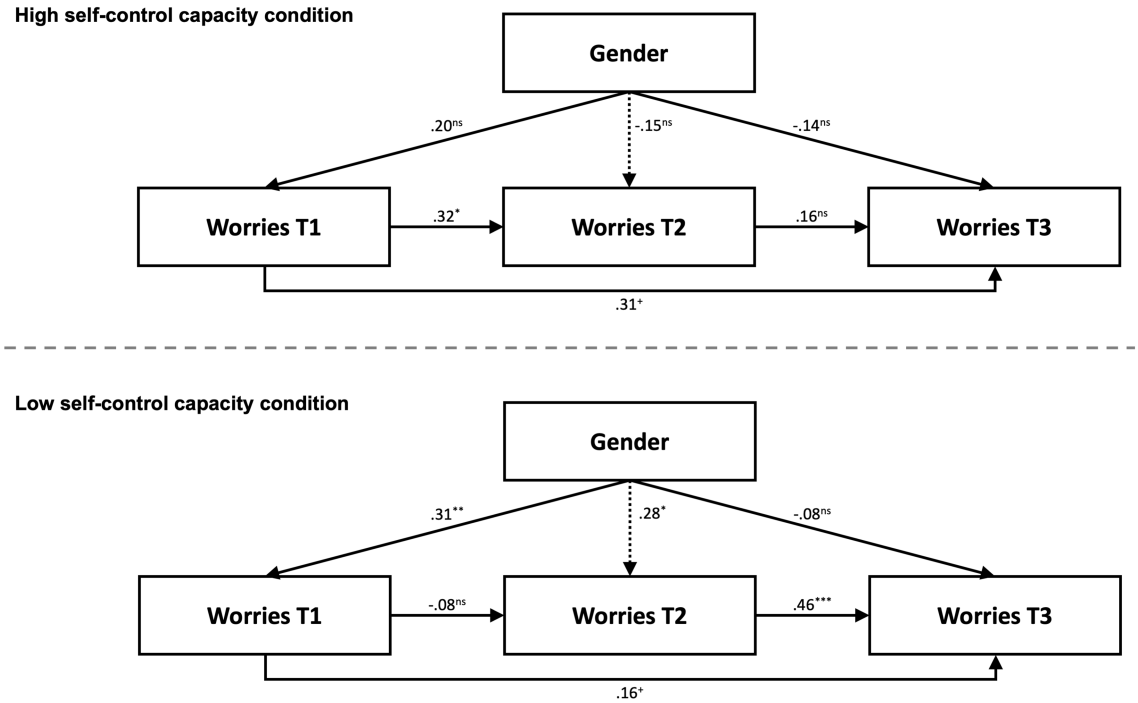


FIGURE 1 | Multiple-group autoregressive path model including gender as a covariate. Top: High self-control capacity condition. Bottom: Low self-control capacity condition. Coding of gender: 0 = male, 1 = female. T1/T2/T3 = first/second/third measurement times. Presented are standardized path coefficients. Dotted lines differ significantly between the two conditions. $N = 102$. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, $^{+}p < 0.10$, ns = nonsignificant.

TABLE 1 | Means and standard deviations (in parentheses) of reported worry at the three times of measurement, separated by self-control capacity condition and gender.

	Self-control capacity high		Self-control capacity low	
Variable	**Females**	**Males**	**Females**	**Males**
Worry at T1	0.38 (0.31)	0.25 (0.33)	0.43 (0.30)	0.25 (0.23)
Worry at T2	0.53 (0.35)	0.59 (0.37)	0.60 (0.31)	0.46 (0.24)
Worry at T3	0.38 (0.34)	0.45 (0.34)	0.39 (0.25)	0.34 (0.31)

N = 102. Worry = Proportion of the number of worry thoughts in the total number of thoughts. T1/T2/T3 = first/second/third measurement times.

these relations among the worry measures by the self-control capacity condition was not significant (**Table 2**).

We also applied multiple-group autoregressive path models to examine whether performance was predicted by gender in different experimental conditions (see the model depicted in **Figure 2**). In contrast to the high self-control capacity condition, there was a significant path from gender to performance at the first measurement time in the low self-control capacity condition (**Figure 2**), suggesting that the performance of female compared with male university students was lower only when self-control capacity was low. However, this path as well as the other paths in the model were not significantly different between the two self-control capacity conditions (**Table 3**). In neither of the two self-control capacity conditions was gender predictive of performance at the second measurement time.

DISCUSSION

Present Findings

The present experiment aimed to show that situational differences in the capacity to exert self-control can determine how strongly stereotype threat is associated with worry in female university students during math test situations. We found some supporting evidence for this prediction; however, the pattern we obtained requires closer examination. In line with our assumption and previous studies on test anxiety (Bertrams et al., 2013), the only significant paths from gender to the extent of worry emerged in the experimental condition in which the self-control capacity was relatively low. In this case, female university students reported more worry than their male counterparts. This is unsurprising, as the female but not the male participants in our study received threatening instruction with respect to the math performance situation. In contrast, we did not find any significant gender–worry relations in the high self-control capacity condition. This finding can be interpreted such that stereotype threat has adverse effects on cognition, as has been theorized and empirically demonstrated in the past (Cadinu et al., 2005; Schmader et al., 2008). However, such stereotype threat effects are controllable and repressible. When their self-control capacity is intact, individuals can deliberately

focus their stream of thought on ongoing tasks, also directing them away from threat-related worry. This effect was mostly pronounced in the middle of the performance situation (i.e., at the second measurement time). For the worry that was reported between the two performance blocks and referred to the first experience with the math tasks, the relation between gender and worry was significantly different for the two self-control capacity conditions.

The experimental conditions did not significantly differ in the relation between gender and worry at the first measurement time, even though the significance and non-significance of this relation in the low and high self-control capacity conditions, respectively, confirmed our expectations. In the high self-control capacity condition, the female university students could initially have experienced some threat that did not build up to a full stereotype threat effect but nevertheless increased the gender–worry relation. Note that at the first measurement time, the students had already been confronted with the stereotype instruction but had not yet seen the math tasks; therefore, uncertainty about what to expect is likely to have contributed to a threat experience. As a result, the difference between the two experimental conditions in the model path at the first measurement time might not have been large enough to reach statistical significance. The influence of uncertainty then vanished after the math tasks had been received, leading to a significant difference in the stereotype threat effect at the second measurement time.

Even in the low self-control capacity condition, the threatened female participants did not experience more worry than the non-threatened male participants at the third measurement time. There are at least two possible explanations for this finding. First, the third thought protocol was placed after the second performance block (i.e., when the performance situation had already ended). Although the instruction for the thought protocol was to report the thoughts during the second math block, it seems reasonable to assume that some worried female participants were relieved about the end of the performance situation. This relief might have biased the third thought protocol. Second, the stereotype threat effect could have declined during the math performance situation due to factors such as

TABLE 2 | Significance of differences in standardized path coefficients between the high and low self-control capacity conditions.

Path	Wald χ^2	*df*	*p*
Gender ➔ Worry T1	0.38	1	0.54
Gender ➔ Worry T2	4.63	1	0.03
Gender ➔ Worry T3	0.12	1	0.73
Worry T1 ➔ Worry T2	3.44	1	0.06
Worry T1 ➔ Worry T3	0.56	1	0.46
Worry T2 ➔ Worry T3	2.15	1	0.14

N = 102. T1/T2/T3 = first/second/third measurement times.

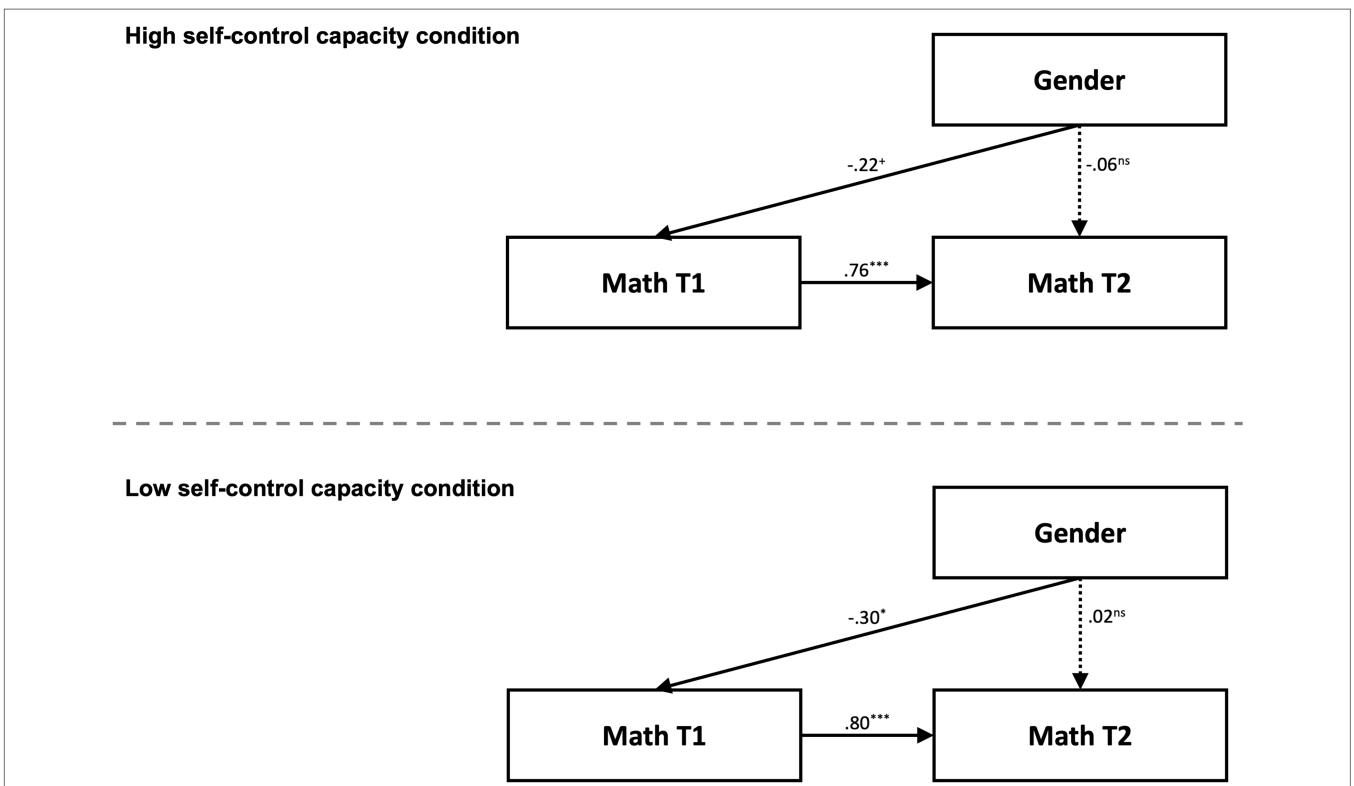


FIGURE 2 | Multiple-group autoregressive path model including gender as a covariate. Top: High self-control capacity condition. Bottom: Low self-control capacity condition. Coding of gender: 0 = male, 1 = female. T1/T2 = first/second measurement times. Presented are standardized path coefficients. Dotted lines differ significantly between the two conditions. $N = 102$. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, $^{+}p < 0.10$, ns = nonsignificant.

TABLE 3 | Significance of differences in standardized path coefficients between the high and low self-control capacity conditions.

Path	Wald χ^2	*df*	*p*
Math T1 ➔ Math T2	0.11	1	0.75
Gender ➔ Math T1	0.15	1	0.70
Gender ➔ Math T2	0.46	1	0.50

N = 102. T1/T2 = first/second measurement times.

habituation. Similarly, in Cadinu et al.'s (2005) study, the relation between stereotype threat and worry during the second half of a math performance situation was relatively low compared with the first half. At the moment, we cannot determine whether the lack of a stereotype threat effect at the third measurement time was caused by a methodological problem or represents a typically occurring pattern.

We obtained an interesting pattern of relations for the stabilities between the worry measures at different measurement times. In the high self-control capacity condition, the worry measures at the first but not the second measurement time predicted worry at the third measurement time. One explanation is that worry at the second measurement time did not incrementally add even more worry beyond the amount of worry at the first measurement time. However, why was worry at the first measurement time unrelated to the worry measures at the other two times in the low self-control capacity condition? Unfortunately, we cannot answer this question satisfactorily, but suspect that a qualitative shift in worry occurred as a result of contact with the math tasks of the first block. Over and above the stereotype threat, some individuals could have been perturbed just by the announcement of a performance situation. Given their momentary lack of self-control capacity to regulate their negative experiences, this might have impacted their reports of worry regarding the period during and immediately after the announcement (i.e., the thought protocol at the first measurement time). After the first confrontation with the performance situation was made (i.e., at the second measurement time), this influence of diffuse expectations might have ceased. Instead, at the second measurement time, the participants' perception of how they were going to actually deal with the performance situation determined the worry. Thus, some individuals may have experienced the actual performance situation as less alarming than expected, and vice versa; others may have seen their expectations as confirmed, overall resulting in the lack of a relation between the worry measures at the first and second measurement times.

Implications

The present findings extend knowledge about the moderating situational circumstances under which stereotype threat can be harmful. Exerting self-control during an unrelated demand can apparently undermine personal resistibility against self-threatening information in subsequent performance contexts. Therefore, it is advisable not to place self-control requirements in advance of important performance situations, particularly for individuals who are at risk of being stereotyped, as we recreated in the present study for females in the context of math. Possibly, similar effects are at work for males in the context of language-related tasks (Bedyńska et al., 2021; but see Chaffee et al., 2020). Viewed from a positive perspective, the present findings also indicate that stereotyping information may not, in any case, be irritating for the individuals concerned. Momentary self-control capacity might thus be a moderator variable, which can explain why there are numerous studies that found stereotype threat effects, while others did not replicate this finding (e.g., Chaffee et al., 2020; McGuire et al., 2021). In general, unconsidered moderator variables have been argued to be one crucial reason for inconsistent findings (e.g., Bertrams et al., 2013; Stroebe and Strack, 2014).

Furthermore, this study complements the research on self-control capacity and test anxiety (Bertrams et al., 2013). It was previously found that the relation between anxiety (instead of stereotype threat) and worry during a performance situation was moderated by momentary self-control capacity. When self-control capacity was relatively intact, trait test anxiety and worry during the performance situation were unrelated; however, when the self-control capacity was reduced by an unrelated previous self-control demand, higher trait test anxiety was associated with a more pronounced experience of distracting worry in the subsequent performance situation. Thus, across different studies, we found a pattern that was largely consistent across various threats (i.e., threat caused by trait test anxiety as a personality variable and threat caused by stereotype activation for a specific group). In this regard, in the present study, trait test anxiety did not differ among gender, self-control capacity condition, or their combination. In summary, there is accumulating evidence that self-control is important in potentially threatening performance contexts.

Limitations and Future Research

Differences in perceived competence that we unexpectedly found in dependence of gender and self-control condition may constitute an alternative explanation for our results. However, when incorporated as covariate, perceived competence did not change the findings. In addition, perceived competence did not predict worry at any time of measurement for any combination of gender and self-control capacity condition. Therefore, we assume that our results were not caused by differences in perceived competence. Still, further in-depth research may shed more light on the role of perceived competence. In this respect, it may be relevant that previous research has usually revealed no effects of the applied self-control capacity manipulation on perceived competence (e.g., Bertrams et al., 2013), but there is also initial evidence that self-control capacity manipulations affected self-efficacy (Graham et al., 2017), a variable that has some conceptual overlap with perceived competence.

In this study, we focused on a direct effect of the presence of a gender stereotype: worry (Schmader et al., 2008). For this reason, we embedded our dependent variable measure into a performance situation with math tasks. The presentation of these tasks was intended to frame the worry measures. Therefore, the math tasks were highly standardized in two blocks with

an unusual response format and applied within a short time limit. In other words, our primary interest was not to find a reliable and valid performance measure. Actually, we think that the applied performance task is not appropriate or useful to draw conclusions about effects on real performance, for instance, because the time limit was much too short to represent a test with sufficient construct validity (see Lu and Sireci, 2007). While the chosen procedure may have been useful for measuring and analyzing the central variable of worry, it also implies a limitation of our study, namely, that we could not determine the impact of stereotype threat on performance. Indeed, no such empirical evidence was found in the present study. In future studies, the effects of a combination of self-control capacity and stereotype threat on performance should be addressed directly. In this regard, we think that measuring performance while thought protocols are embedded would not be ideal, as the interruptions may have uncontrolled effects on performance measurement. Instead, we recommend using a multistep experimental approach to detect a causal chain (Spencer et al., 2005).

Given the means of the worry measures, it cannot be ruled out that self-control capacity moderated a stereotype lift effect (Walton and Cohen, 2003), rather than a stereotype threat effect. Possibly, the stereotype threat instruction we used may have increased confidence and reduced worry in the male participants. At the moment, however, we cannot provide a theoretical basis beyond pure speculation as to why low self-control capacity would facilitate the stereotype lift effect. Basically, the assumption that low compared to high self-control capacity would be associated with lower worry during test situations contradicts previous reasoning (e.g., Bertrams et al., 2013). The potential interplay between self-control capacity and stereotype lift could still be a promising subject for future examination.

The present research can also be challenged by the ongoing debate on whether individuals' self-control capacity is actually reduced after initial self-control exertion. Some authors deny the view that initial self-control detrimentally affects subsequent self-control and cognition (e.g., Carter et al., 2015), while others defend it (e.g., Baumeister et al., 2020; see also Englert and Bertrams, 2021). We believe that discussing and investigating the underlying mechanisms and potential moderator variables regarding situational fluctuations in self-control capacity is an appropriate way forward (e.g., Bertrams, 2020, 2021). In the present study, we found an effect of our manipulation of self-control capacity on worry; however, there might be different ways to interpret it. For instance, the manipulation might have undermined self-control by exhausting a self-regulatory resource (Baumeister et al., 2007) or by activating the cognitive concepts of fatigue and energy saving (Bertrams, 2020). Future research can determine in more detail when and how self-control exertion can harm subsequent self-control. These insights can advance the understanding of our findings and how interventions based on them can be optimally designed.

To conclude, the results from the present study represent a step toward understanding the conditions of females' underachievement in math under stereotype threat. By explicitly showing the moderating function of students' momentary self-control capacity in the relation between stereotype threat and worry, our study contributes to a deeper understanding of the mechanisms of stereotype threat.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AB developed the study concept and design and collected the data. AB and JR analyzed and interpreted the data. AB, CL, FM, and JR wrote the manuscript. All authors approve the final version to be published, contributed meaningfully to the paper, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

Open access funding provided by University of Bern.

REFERENCES

Baumeister, R. F., and Vohs, K. D. (2016). "Strength model of self-regulation as limited resource: assessment, controversies, update," in *Advances in Experimental Social Psychology*. eds. J. M. Olson and M. P. Zanna (United States: Elsevier Academic Press), 67–127.

Baumeister, R. F., Vohs, K. D., and Tice, D. M. (2007). The strength model of self-control. *Curr. Dir. Psychol. Sci.* 16, 351–355. doi: 10.1111/j.1467-8721.2007.00534.x

Baumeister, R. F., Vonasch, A. J., and Sjåstad, H. (2020). "The long reach of self-control," in *Surrounding Self-Control*. ed. A. R. Mele (England: Oxford University Press).

Bedyńska, S., Rycielski, P., and Jabłońska, M. (2021). Measuring stereotype threat at math and language arts in secondary school: validation of a questionnaire. *Front. Psychol.* 12:553964. doi: 10.3389/fpsyg.2021.553964

Beilock, S. L., Gunderson, E. A., Ramirez, G., and Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1860–1863. doi: 10.1073/pnas.0910967107

Beilock, S. L., Rydell, R. J., and McConnell, A. R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spillover. *J. Exp. Psychol. Gen.* 136, 256–276. doi: 10.1037/0096-3445.136.2.256

Bertrams, A. (2020). A schema-activation approach to failure and success in self-control. *Front. Psychol.* 11:2256. doi: 10.3389/fpsyg.2020.02256

Bertrams, A. (2021). The cognitive association between effortful self-control and decreased vitality. *Front. Psychol.* 12:631914. doi: 10.3389/fpsyg.2021.631914

Bertrams, A., Baumeister, R. F., Englert, C., and Furley, P. (2015). Ego depletion in color priming research: self-control strength moderates the detrimental effect of red on cognitive test performance. *Personal. Soc. Psychol. Bull.* 41, 311–322. doi: 10.1177/0146167214564968

Bertrams, A., and Englert, C. (2014). Test anxiety, self-control, and knowledge retrieval in secondary school students. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 46, 165–170. doi: 10.1026/0049-8637/a000111

Bertrams, A., Englert, C., and Dickhäuser, O. (2010). Self-control strength in the relation between trait test anxiety and state anxiety. *J. Res. Pers.* 44, 738–741. doi: 10.1016/j.jrp.2010.09.005

Bertrams, A., Englert, C., Dickhäuser, O., and Baumeister, R. F. (2013). Role of self-control strength in the relation between anxiety and cognitive performance. *Emotion* 13, 668–680. doi: 10.1037/a0031921

Brodish, A. B., and Devine, P. G. (2009). The role of performance–avoidance goals and worry in mediating the relationship between stereotype threat and performance. *J. Exp. Soc. Psychol.* 45, 180–185. doi: 10.1016/j.jesp.2008.08.005

Cadinu, M., Maass, A., Rosabianca, A., and Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychol. Sci.* 16, 572–578. doi: 10.1111/j.0956-7976.2005.01577.x

Carter, E. C., Kofler, L. M., Forster, D. E., and McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: self-control does not seem to rely on a limited resource. *J. Exp. Psychol. Gen.* 144, 796–815. doi: 10.1037/xge0000083

Chaffee, K. E., Lou, N. M., and Noels, K. A. (2020). Does stereotype threat affect men in language domains? *Front. Psychol.* 11:1302. doi: 10.3389/fpsyg.2020.01302

Chalabaev, A., Major, B., Sarrazin, P., and Cury, F. (2012). When avoiding failure improves performance: stereotype threat and the impact of performance goals. *Motiv. Emot.* 36, 130–142. doi: 10.1007/s11031-011-9241-x

Dang, J., Barker, P., Baumert, A., Bentvelzen, M., Berkman, E., Buchholz, N., et al. (2020). A multilab replication of the ego depletion effect. *Soc. Psychol. Personal. Sci.* 12, 14–24. doi: 10.1177/1948550619887702

Désert, M., Préaux, M., and Jund, R. (2009). So young and already victims of stereotype threat: socio-economic status and performance of 6 to 9 years old children on Raven's progressive matrices. *Eur. J. Psychol. Educ.* 24, 207–218. doi: 10.1007/BF03173012

Dummel, S., and Rummel, J. (2016). Effects of ego-depletion on choice behaviour in a multi-attribute decision task. *J. Cogn. Psychol.* 28, 374–383. doi: 10.1080/20445911.2015.1135929

Else-Quest, N. M., Hyde, J. S., and Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychol. Bull.* 136, 103–127. doi: 10.1037/a0018053

Englert, C., and Bertrams, A. (2017). Ego depletion negatively affects knowledge retrieval in secondary school students. *Educ. Psychol.* 37, 1057–1066. doi: 10.1080/01443410.2017.1313963

Englert, C., and Bertrams, A. (2021). Again, No evidence for or against the existence of ego depletion: opinion on "A multi-site preregistered paradigmatic test of the ego depletion effect". *Front. Hum. Neurosci.* 15:658890. doi: 10.3389/fnhum.2021.658890

Evans, A. B., Copping, K., Rowley, S. J., and Kurtz-Costes, B. (2011). Academic self-concept in black adolescents: do race and gender stereotypes matter? *Self Identity* 10, 263–277. doi: 10.1080/15298868.2010.485358

Garrison, K. E., Finley, A. J., and Schmeichel, B. J. (2019). Ego depletion reduces attention control: evidence from two high-powered preregistered experiments. *Personal. Soc. Psychol. Bull.* 45, 728–739. doi: 10.1177/0146167218796473

Gerstenberg, F. X. R., Imhoff, R., and Schmitt, M. (2012). 'Women are bad at math, but I'm not, am I?' Fragile mathematical self-concept predicts vulnerability to a stereotype threat effect on mathematical performance. *Eur. J. Personal.* 26, 588–599. doi: 10.1002/per.1836

Graham, J. D., Martin Ginis, K. A., and Bray, S. R. (2017). Exertion of self-control increases fatigue, reduces task self-efficacy, and impairs performance of resistance exercise. *Sport Exerc. Perform. Psychol.* 6, 70–88. doi: 10.1037/spy0000074

Hedges, L. V., and Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science* 269, 41–45. doi: 10.1126/science.7604277

Hofmann, W., Schmeichel, B. J., and Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends Cogn. Sci.* 16, 174–180. doi: 10.1016/j.tics.2012.01.006

Hyde, J. S. (2005). The gender similarities hypothesis. *Am. Psychol.* 60, 581–592. doi: 10.1037/0003-066X.60.6.581

Job, V., Dweck, C. S., and Walton, G. M. (2010). Ego depletion-is it all in your head? Implicit theories about willpower affect self-regulation. *Psychol. Sci.* 21, 1686–1693. doi: 10.1177/0956797610384745

Kurzban, R., Duckworth, A., Kable, J. W., and Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behav. Brain Sci.* 36, 661–679. doi: 10.1017/S0140525X12003196

Lamont, R. A., Swift, H. J., and Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: negative stereotypes, not facts, do the damage. *Psychol. Aging* 30, 180–193. doi: 10.1037/a0038586

Lindberg, S. M., Hyde, J. S., Petersen, J. L., and Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychol. Bull.* 136, 1123–1135. doi: 10.1037/a0021276

Lindner, M. A., Lüdtke, O., and Nagy, G. (2019). The onset of rapid-guessing behavior over the course of testing time: A matter of motivation and cognitive resources. *Front. Psychol.* 10, 1–15. doi: 10.3389/fpsyg.2019.01533

Lindner, C., Nagy, G., Ramos Arhuis, W. A., and Retelsdorf, J. (2017). A new perspective on the interplay between self-control and cognitive performance: modeling progressive depletion patterns. *PLoS One* 12:e0180149. doi: 10.1371/journal.pone.0180149

Lindner, C., and Retelsdorf, J. (2020). Perceived — and not manipulated — self-control depletion predicts students' achievement outcomes in foreign language assessments. *Educ. Psychol.* 40, 490–508. doi: 10.1080/01443410.2019.1661975

Lu, Y., and Sireci, S. G. (2007). Validity issues in test speededness. *Educ. Meas. Issues Pract.* 26, 29–37. doi: 10.1111/j.1745-3992.2007.00106.x

Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., and Rodgers, B. (1999). A short form of the positive and negative affect schedule: evaluation of factorial validity and invariance across demographic variables in a community sample. *Personal. Individ. Differ.* 27, 405–416. doi: 10.1016/S0191-8869(98)00251-7

Martinot, D., Bagès, C., and Désert, M. (2012). French children's awareness of gender stereotypes about mathematics and reading: when girls improve their reputation in math. *Sex Roles* 66, 210–219. doi: 10.1007/s11199-011-0032-3

McGuire, L., Monzavi, T., Hoffman, A. J., Law, F., Irvin, M. J., Winterbottom, M., et al. (2021). Science and math interest and gender stereotypes: The role of educator gender in informal science learning sites. *Front. Psychol.* 12:503237. doi: 10.3389/fpsyg.2021.503237

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Muthén, L. K., and Muthén, B. O. (2017). *Mplus: Statistical Analysis with Latent Variables: User's Guide (Version 8)*. New York: Wiley.

Nosek, B. A., Banaji, M. R., and Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *J. Pers. Soc. Psychol.* 83, 44–59. doi: 10.1037//0022-3514.83.1.44

Nussbaum, A. D., and Steele, C. M. (2007). Situational disengagement and persistence in the face of adversity. *J. Exp. Soc. Psychol.* 43, 127–134. doi: 10.1016/j.jesp.2005.12.007

Passolunghi, M. C., Ferreira, T. I. R., and Tomasetto, C. (2014). Math–gender stereotypes and math-related beliefs in childhood and early adolescence. *Learn. Individ. Differ.* 34, 70–76. doi: 10.1016/j.lindif.2014.05.005

Pennington, C. R., Heim, D., Levy, A. R., and Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLoS One* 11:e0146487. doi: 10.1371/journal.pone.0146487

Plante, I., Théorêt, M., and Favreau, O. E. (2009). Student gender stereotypes: contrasting the perceived maleness and femaleness of mathematics and language. *Educ. Psychol.* 29, 385–405. doi: 10.1080/01443410902971500

Reilly, D., Neumann, D. L., and Andrews, G. (2015). Sex differences in mathematics and science achievement: a meta-analysis of National Assessment of educational Progress assessments. *J. Educ. Psychol.* 107, 645–662. doi: 10.1037/edu0000012

Reilly, D., Neumann, D. L., and Andrews, G. (2019). Investigating gender differences in mathematics and science: results from the 2011 trends in mathematics and science survey. *Res. Sci. Educ.* 49, 25–50. doi: 10.1007/s11165-017-9630-6

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *J. Exp. Soc. Psychol.* 38, 194–201. doi: 10.1006/jesp.2001.1500

Schmader, T., and Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *J. Pers. Soc. Psychol.* 85, 440–452. doi: 10.1037/0022-3514.85.3.440

Schmader, T., Johns, M., and Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychol. Rev.* 115, 336–356. doi: 10.1037/0033-295X.115.2.336

Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *J. Exp. Psychol. Gen.* 136, 241–255. doi: 10.1037/0096-3445.136.2.241

Schmeichel, B. J., Vohs, K. D., and Baumeister, R. F. (2003). Intellectual performance and ego deple- tion: role of the self in logical reasoning and other information processing. *J. Pers. Soc. Psychol.* 85, 33–46. doi: 10.1037/0022-3514.85.1.33

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2022). Abiturnoten im Ländervergleich [Comparison of Abitur grades among the federal states]. Available at: https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html (Accessed January 7, 2022).

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366. doi: 10.1177/0956797611417632

Spencer, S. J., Logel, C., and Davies, P. G. (2016). Stereotype threat. *Annu. Rev. Psychol.* 67, 415–437. doi: 10.1146/annurev-psych-073115-103235

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *J. Exp. Soc. Psychol.* 35, 4–28. doi: 10.1006/jesp.1998.1373

Spencer, S. J., Zanna, M. P., and Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *J. Pers. Soc. Psychol.* 89, 845–851. doi: 10.1037/0022-3514.89.6.845

Spielberger, C. D., and Vagg, P. R. (1995). "Test anxiety: a transactional process model," in *Test Anxiety: Theory, Assessment, and Treatment*. eds. C. D. Spielberger and P. R. Vagg (United States: Taylor & Francis), 3–14.

Steele, C. M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *Am. Psychol.* 52, 613–629. doi: 10.1037/0003-066X.52.6.613

Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J. Pers. Soc. Psychol.* 69, 797–811. doi: 10.1037/0022-3514.69.5.797

Stroebe, W., and Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspect. Psychol. Sci.* 9, 59–71. doi: 10.1177/1745691613514450

Tempel, T., and Neumann, R. (2014). Stereotype threat, test anxiety, and mathematics performance. *Soc. Psychol. Educ.* 17, 491–501. doi: 10.1007/s11218-014-9263-9

Tomasetto, C. (2019). "Gender stereotypes, anxiety, and math outcomes in adults and children," in *Mathematics Anxiety: What Is Known, and What Is Still Missing*. eds. I. C. Mammarella, S. Caviola and A. Dowker (London: Routledge).

Tomasetto, C., Alparone, F. R., and Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Dev. Psychol.* 47, 943–949. doi: 10.1037/a0024047

Vandello, J. A., Bosson, J. K., Cohen, D., Burnaford, R. M., and Weaver, J. R. (2008). Precarious manhood. *J. Pers. Soc. Psychol.* 95, 1325–1339. doi: 10.1037/a0012453

Wacker, A., Jaunzeme, J., and Jaksztat, S. (2008). Eine Kurzform des Prüfungsängstlichkeitsinventars TAI-G [A short form of the test anxiety inventory TAI-G]. *Zeitschrift für Pädagogische Psychologie* 22, 73–81. doi: 10.1024/1010-0652.22.1.73

Walton, G. M., and Cohen, G. L. (2003). Stereotype lift. *J. Exp. Soc. Psychol.* 39, 456–467. doi: 10.1016/S0022-1031(03)00019-2

Wiesner, C. D., and Lindner, C. (2017). Weakening self-control biases the emotional evaluation of appetitive cues. *PLoS One* 12:e0170245. doi: 10.1371/journal.pone.0170245

Yeung, N. C. J., and von Hippel, C. (2008). Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. *Accid. Anal. Prev.* 40, 667–674. doi: 10.1016/j.aap.2007.09.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fpsyg.2022.820254

Exploring the Nature of Teachers' Math-Gender Stereotypes: The Math-Gender Misconception Questionnaire

Anna-Sophia Dersch[1], Anke Heyder[2]* and Alexander Eitel[1]*

[1] Teaching and Learning With Media, Department of Educational Psychology, Institute for Psychology, Justus Liebig University Giessen, Giessen, Germany, [2] Department of Educational and Differential Psychology, Institute for Psychology, TU Dortmund University, Dortmund, Germany

Stereotypes of girls having weaker mathematical abilities than boys (math-gender stereotypes) are one factor reducing women's representation in mathematics. Teachers, as powerful socializers, often hold math-gender stereotypes. Reducing math-gender stereotypes in (student) teachers thus may foster women's representation in mathematics. Yet knowing the stereotypes' underlying assumptions is crucial to reducing it. Do math-gender stereotypes reflect elaborate, disproven theories about gender differences in math, meaning *math-gender misconceptions*? And if so, which math-gender misconceptions are behind math-gender stereotypes? This is the focus of the present research. The relevant literature implies the existence of three distinct misconceptions: (1) *empathizing-systemizing* ("As girls think rather empathically and boys think rather systematically, boys are on average more talented in math than girls"), (2) *girls' compensation* ("To achieve equally good grades in mathematics, boys have to make less effort because they are more talented than girls are"), and (3) *girls' non-compensability* ("Despite their on average stronger effort, girls are normally less proficient in math than boys"). We assessed these misconceptions in a student teacher sample (N = 303) using our newly developed *Math-Gender Misconceptions Questionnaire*. Our results offer support for the expected three-factor structure of math-gender misconceptions. All three math-gender misconceptions showed good to acceptable scale reliabilities. On average, preservice teachers did not hold (strong) math-gender misconceptions. But a subgroup of 48.2% of preservice teachers held at least one of the three misconceptions. The *empathizing-systemizing* misconception was the most prevalent (32.0%) among the three misconceptions. Descriptively, endorsing the math-gender stereotype correlated most strongly with the *empathizing-systemizing* (r = 0.43) and the *girls' compensation misconception* (r = 0.44). This may indicate that especially these two misconceptions partly underlie math-gender stereotypes. As a consequence, refutation instructions designed to reduce these misconceptions may be a promising method to weaken math-gender stereotypes. Further research is needed to investigate to what degree reducing the present misconceptions is related to reducing math-gender stereotypes. Hence, this study is the first one of a planned series of studies on the relation between math-gender misconceptions and math-gender stereotypes.

Keywords: stem education, misconceptions, questionnaire, teacher education, math-gender stereotypes

OPEN ACCESS

Edited by:
Jan Retelsdorf,
University of Hamburg, Germany

Reviewed by:
Joan Barth,
University of Alabama, United States
David Reilly,
Griffith University, Australia

****Correspondence:***
Anna-Sophia Dersch
anna-sophia.dersch@psychol.uni-giessen.de
Anke Heyder
anke.heyder@tu-dortmund.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *22 November 2021*
Accepted: *21 March 2022*
Published: *14 April 2022*

Citation:
Dersch A-S, Heyder A and Eitel A (2022) Exploring the Nature of Teachers' Math-Gender Stereotypes: The Math-Gender Misconception Questionnaire. Front. Psychol. 13:820254. doi: 10.3389/fpsyg.2022.820254

INTRODUCTION

Stereotypes of girls having weaker mathematical abilities than boys (math-gender stereotypes) are widely prevalent in Western societies (Nosek et al., 2010). Math-gender stereotypes reduce girls' interest, motivation, and performance in math, and lead to women being less likely to pursue mathematical professions (e.g., Wang and Degol, 2017). Teachers, as powerful socializers, also endorse math-gender stereotypes (e.g., Gunderson et al., 2012). Reducing math-gender stereotypes in (student) teachers thus seems a promising way to foster the representation of girls and women in mathematics. However, to address these stereotypes effectively, we must know about their nature and underlying assumptions. Do math-gender stereotypes reflect elaborate, yet disproven, theories about gender differences in mathematical abilities, that is, *misconceptions* (e.g., Eitel et al., 2021)? And if so, which misconceptions about mathematical abilities exist? We aim to answer these questions in the present research. This is important, because such misconceptions do not dissipate on their own – instead, overcoming them requires specific instructions in teacher education (refutation texts; Eitel et al., 2019; Menz et al., 2021).

In the literature, we identified three potential misconceptions associated with math-gender stereotypes about mathematical abilities: First, boys are assumed to be inherently better in math, because they supposedly think more systematically than girls, whereas girls think more empathically (Baron-Cohen, 2005; for disprove, see Escovar et al., 2016). Secondly, girls are assumed to succeed as well as boys in math only because they are hardworking, whereas boys are simply talented. This belief was detected in teachers and other socializers (Tiedemann, 2002; Robinson-Cimpian et al., 2014; Sáinz et al., 2020). Thirdly, if mathematical abilities are perceived as fixed (for theories of *fixed* and *growth mindset* and their influence on learners, see Dweck, 1999; Gunderson et al., 2017) and girls are ascribed less mathematical talent, then girls would be unable to compensate for their poorer mathematical abilities. In this study, we developed the *Math-Gender Misconception Questionnaire* (*MGMQ*) to investigate to what degree these three potential misconceptions are empirically separable, present in a student teacher sample, and linked to related constructs such as fixed mindsets of math ability (e.g., Leslie et al., 2015). Note that this study is the first of a planned series of studies on the relation between math-gender misconceptions and math-gender stereotypes.

Gender Stereotypes About Mathematical Abilities

There is evidence that girls' – and boys' mathematical abilities are not inherently different (Lachance and Mazzocco, 2006; Kersey et al., 2019). However, with age, math-gender differences favoring male students emerge in some countries (Else-Quest et al., 2010). These gender differences are relatively small compared to other performance differences (e.g., caused by economic status; Bloom et al., 2008). Further, such differences are usually found in older learners (e.g., Reilly et al., 2015) already influenced by societal gender attitudes (Eliot, 2010). Accordingly, gender differences are mediated by sex-role identity and related to cultural opportunity structures for women (Reilly, 2012). Moreover, gender stereotypes about girls' and women's lesser abilities in science, technology, engineering, and math (STEM) are widely prevalent in Western cultures (Nosek et al., 2010; Nosek and Smyth, 2011; Hand et al., 2017) and predict women's lower STEM engagement (Hyde et al., 1990; Halpern et al., 2007; see Nosek and Smyth, 2011; for similar findings on reading and boys, see e.g., Retelsdorf et al., 2015; Muntoni and Retelsdorf, 2018). In this vein, in many Western countries, women remain underrepresented in the mathematical professions (Wang and Degol, 2017). The societal stereotypes of girls' and women's lesser math abilities (*math-gender stereotypes*) influence children from an early age (e.g., Eliot, 2010). Math-gender stereotypes are conveyed by parents, peers, and teachers (see e.g., Hannover, 2008). As school is especially important for children's socialization (Wentzel, 2014), children are prone to being influenced by teachers' math-gender stereotypes. According to the *Model of Achievement Related Choices* (Eccles et al., 1983), teachers, as part of the cultural milieu, hold gender stereotypes including math-gender stereotypes (Eccles, 2011; Gunderson et al., 2012). Teachers have more positive attitudes about male students' math performance, overrate male students' mathematical abilities and have higher expectations regarding male students' mathematical success (Riegle-Crumb and Humphries, 2012; Robinson-Cimpian et al., 2014; for a literature review, see Li, 1999). Further, teachers attribute failure in math to a lack of talent among girls, but to a lack of effort among boys (Tiedemann, 2002), and demonstrate a gender bias when evaluating students' performance in an experimental setting (underrating equal performance outcomes if they assume female learners achieved them (Avitzour et al., 2020; see also Holder and Kessels, 2017). Furthermore, the Eccles et al. (1983) model proposes that teachers' beliefs and behaviors influence their students' own gender roles and stereotypes (see Eccles, 2011). Teachers' own math-gender stereotypes thus predict students' math-gender stereotypes (Keller, 2001). Although these math-gender stereotypes seem to have decreased in school children (e.g., Passolunghi et al., 2014), recent research still suggests that even primary school children hold the perception of math being male-typed (Miller et al., 2015). These stereotypes then influence students' expectation of success and subjective task value (e.g., in mathematics), which in turn influences students' achievement-related choices. Math-gender stereotypes of girls lead to girls tending to make academic choices against mathematics (see Eccles, 2011). Apart from academic choices, math-gender stereotypes influence girls' sense of identity. The idea of math being male-typed (Miller et al., 2015) leads to girls developing less interest or preference for math when forming their identity. Thus, girls do not engage further with math, as girls try to establish their identity as distinct from the boys' identity and from male-typed interests (Bian et al., 2017). All in all, math-gender stereotypes reduce girls' interest, motivation, and performance in math, and, ultimately, lead to women being less likely to pursue mathematical professions (e.g., Wang and Degol, 2017). Further, according to learning theories, girls (and

boys) learn to behave according to gender stereotypes because parents, *teachers* and peers reinforce them for doing so (Mischel, 1966; Hannover, 2008). This process of operant conditioning leads to girls' engaging less with math as teachers – due to their math-gender stereotype – reinforce girls less than they reinforce boys for engaging with math. Besides that, math-gender stereotypes influence girls' – and women's performance through social-psychological mechanisms such as *self-fulfilling prophecies* or *stereotype threat*. When societal stereotypes are activated, girls are more likely to behave in a way that fulfills societal stereotypes and expectations. For example, teachers implicitly expressing their math-gender stereotypes and thus treating girls differently may instigate a worse math performance [see *self-fulfilling prophecy* (Merton and Merton, 1968)]. Just the fear itself of negative judgment in light of the math-gender stereotype can cause a disruption leading to girls' performing worse in math [see *stereotype threat* (Steele and Aronson, 1995)]. This means that teachers, who – because of their math-gender stereotypes – expect girls to perform worse, in fact contribute to female learners actually performing worse in standardized math tests (Geis, 1993; Smith et al., 1999; Spencer et al., 1999).

Finally, as powerful socializers, teachers do not only endorse math-gender stereotypes, their math-gender stereotypes influence girls' math attitudes and performance negatively (Gunderson et al., 2012; Carlana, 2019). Reducing teachers' stereotypes may therefore represent a means to increase women's representation in mathematics. To weaken stereotypes, however, it is important to know about their nature and underlying assumptions, which is in the focus of this research.

Interrelation of Math-Gender Stereotypes and Math-Gender Misconceptions

Math-gender stereotypes and misconceptions about math abilities based on gender (*math-gender misconceptions*) are two theoretically related but separable constructs.

Stereotypes are based on oversimplified, overgeneralized beliefs (Klineberg, 1951); for instance, beliefs that a certain group member has certain attributes because they are a member of a group (Greenwald et al., 2002). Thus, the math-gender stereotype is the over-simplified, overgeneralized belief of girls having weaker mathematical abilities because of their gender (Math-gender). Stereotypes are rarely fully refuted (FitzGerald et al., 2019; Kollmayer et al., 2020). This may be the case, because the specific reasoning or (mis-)conceptions behind a global stereotype are hard to grasp and therefore hard to target (e.g., by refutation texts; Tippett, 2010). Likewise, empirical evidence showing that math-gender stereotypes persist despite being incorrect (e.g., Gunderson et al., 2012) is paralleled by the scarcity of research on how instruction can overcome these math-gender stereotypes (Kollmayer et al., 2020). In this study, we want to explore the specific reasoning behind teachers' math-gender stereotypes to prospectively provide refutation instruction. More specifically, we want to know whether endorsing math-gender stereotypes is related to holding *math-gender misconceptions* – subjectively plausible, yet disproven, theories about gender differences in mathematical abilities (for misconception definition, see Vosniadou, 1994; Chi and Roscoe, 2002; Hughes et al., 2013).

Math-Gender Misconceptions

Previous research suggests the potential presence of *three* specific misconceptions underlying gender stereotypes about mathematical abilities.

The first potential misconception refers to the Empathizing-Systemizing theory (Baron-Cohen, 2005) to explain the assumption of boys' better inherent mathematical abilities compared to girls' inherent mathematical abilities. The prominent Empathizing-Systemizing theory assumes that biological determinants explain gender differences in math. The Empathizing-Systemizing theory states that, because pre-natal testosterone-exposure is higher in the male fetus than the female, boys develop more systematic thinking in relation to less empathic thinking. Because pre-natal testosterone-exposure is lower in girls than boys, girls develop less systematic thinking in relation to more empathic thinking. According to the Empathizing-Systemizing theory, girls' weaker systematic thinking leads to lower mathematical abilities (Baron-Cohen, 2005). This view, however, is very one-sided and excludes societal factors scientifically proven to be important (e.g., Hannover, 2008; Eliot, 2010; Eccles, 2011; Wang and Degol, 2017). Further, even though female participants in some research did exhibit a higher ratio of empathic to systematic thinking than did men and vice-versa (e.g., Greenberg et al., 2018), this ratio-difference did not predict mathematical performance, even when researched in a huge sample (Escovar et al., 2016). In addition, the idea of empathic thinking being negatively associated with systematic thinking is not very convincing, considering that both refer to the construct of general thinking abilities [general intelligence (g); Gottfredson, 1998]. Consequently, the Empathizing-Systemizing theory itself represents a math-gender misconception (*empathizing-systemizing* misconception).

The second potential misconception, termed *girls' compensation*, refers to the belief that girls achieve similar math results as boys because they are hardworking, whereas boys are simply talented. However, girls actually report *less* intrinsic motivation in math than boys (e.g., Skaalvik and Rankin, 1994; Rodriguez et al., 2020; Heyder et al., 2020). As motivation is a strong predictor for effort and persistence (Skaalvik et al., 2015), girls are likely to be less driven to succeed in math. Girls are therefore very unlikely to achieve similar math results as boys only because they work harder. Furthermore, results from various studies suggest a similar level of mathematical talent in boys and girls: At a young age, girls and boys reveal gender similarities – rather than differences – in neural functioning when engaging with mathematical content (Kersey et al., 2019). In a longitudinal observation of primary school children (Lachance and Mazzocco, 2006), sex differences in math performance measured via standardized tests were minimal to non-existent. These empirical results offer no support for the idea that girls have lower math abilities overall. *Girls' compensation* thus counts as a math-gender misconception.

The third potential misconception, termed *girls' non-compensability*, also refers to the belief about gender differences

TABLE 1 | Descriptive values for misconception items and scale reliabilities of the MGMQ.

Empathizing-systemizing (ES): ω = 0.88; asymptotic ω = 0.90	**Agreement rates (min. = 0, max. = 1)**	**Response certainty (min. = 0, max. = 4)**	**Misconception score[a] (min. = −4, max. = +4)**	**Item-total correlation (min. = 0, max. = 1)**
ES1: *As girls think rather empathically and boys think rather systematically, boys are on average more talented in math than girls*	0.32	2.50 (0.94)	−1.30 (2.33)	0.57
ES2: *Mathematical relationships are usually easier to understand for boys than girls, because boys think in more systematic contexts*	0.39	2.31 (0.92)	−0.65 (2.40)	0.74
ES3: *As boy, more likely think in systematic categories, they fulfill more cognitive prerequisites for math than girls do*	0.39	2.22 (1.00)	−0.73 (2.33)	0.75
ES4: *Female empathy makes it easier for girls to deal with people, while boys are usually more gifted in systematic thinking and thus in math*	0.49	2.50 (0.97)	−0.28 (2.67)	0.77
ES5*: On average, girls think more empathically than boys do, while boys are more talented in systematic thinking and thus also in math*	0.44	2.34 (0.99)	−0.53 (2.49)	0.81
Girls' compensation (GC): ω = 0.76; asymptotic ω = 0.91				
GC1: *Mathematical content often comes easily to boys, while girls on average have to make more effort*	0.14	2.76 (0.86)	−2.16 (1.93)	0.58
GC2: *Girls normally have to work harder to perform as well in math as boys*	0.23	2.63 (0.83)	−1.60 (2.25)	0.61
GC3: *Girls compensate for their usually less aptitude in math compared to boys by being more diligent*	0.48	2.36 (0.91)	−0.19 (2.52)	0.46
GC4: *Girls usually need additional help to perform on par with boys in math*	0.14	2.61 (0.98)	−1.96 (1.98)	0.54
GC5: *To achieve equally good grades in math, boys have to make less effort because they are more talented than girls are*	0.17	2.67 (0.98)	−1.97 (2.05)	0.71
Girls' non-compensability (GN): ω = 0.72; asymptotic ω = 0.68				
GN1: *Since girls are on average less mathematically gifted, they should be assessed with different criteria than boys*	0.05	3.34 (0.87)	−3.10 (1.53)	0.56
GN2: *Girls should be rewarded with good grades for their stronger efforts in math, as they are not naturally as good at math as boys*	0.08	3.08 (0.98)	−2.74 (1.71)	0.62
GN3: *If the top of the class in math is a boy, it is because, in addition to his effort, he possesses a natural talent in math that diligent girls often lack*	0.18	2.80 (1.01)	−2.08 (2.14)	0.47
GN4: *Girls cannot fully compensate for their lack of aptitude for math with their on average greater diligence*	0.14	2.72 (0.89)	−2.11 (1.93)	0.45
GN5: *Despite their on average stronger effort, girls are normally less proficient in math than boys*	0.21	2.56 (0.97)	−1.67 (2.17)	0.43
All items: ω = 0.82; asymptotic ω = 0.69				

Agreement rates represent the proportion of participants agreeing statement. Descriptive values for response certainty and misconception scores represent means and standard deviations (in parentheses).

[a]*Calculated by converting agreement into +1 and disagreement into −1, then multiplied with response certainty.*

in mathematical talent. However, here the focus is on innate differences in mathematical talent that girls cannot compensate for later in life, because talent is assumed to be fixed. This *fixed mindset* is especially common in mathematics and other STEM subjects (e.g., Leslie et al., 2015; Gunderson et al., 2017; Canning et al., 2019) and also identified among teachers (Heyder et al., 2020). A fixed mindset stands in opposition to evidence of educational achievement, such as the *growth mindset* proposed by Dweck (1999, 2015). Accordingly, rather than being fixed, skills can improve over time with practice. However, people who hold the *girls' non-compensability* misconception assume that talent is fixed, and simultaneously ascribe girls less mathematical talent. In so doing, they assume girls cannot compensate for inherent talent differences in mathematical abilities. However, as described before, there is no evidence supporting the idea of girls having lower innate math abilities. Furthermore, the combination of a fixed mindset and lack-of-talent assumptions is especially detrimental for female students' math-attitudes (Dweck, 2015; Heyder et al., 2019, 2020; Muenks et al., 2020) and for their performance (Canning et al., 2021).

Current Study and Hypotheses

In this study, we present the newly developed *Math-Gender Misconception Questionnaire (MGMQ)* to assess teachers' misconceptions about gender differences in mathematics abilities. These misconceptions may underlie stereotypical thinking and behavior (see section "Interrelation of Math-Gender Stereotypes and Math-Gender Misconceptions"). By means of this questionnaire, we investigated to what degree the three potential misconceptions (*empathizing-systemizing*, *girls' compensation*, *girls' non-compensability*) are (1) empirically separable (*structure hypothesis*) and measurable by reliable scales, (2) present in a student teacher sample (*prevalence*

hypothesis), and (3) linked to theoretically related constructs (*association hypothesis*).

Structure Hypothesis

We expect the MGMQ to assess three empirically separable, yet positively interrelated misconceptions. All three of the previously described misconceptions (see section "Interrelation of Math-Gender Stereotypes and Math-Gender Misconceptions") are related to beliefs about gender differences in mathematical talent. Nevertheless, each misconception focuses on a different aspect: The *empathizing-systemizing* misconception provides an over-simplified explanation for the existence of gender differences in mathematical talent. The *girls' compensation* misconception refers to girls managing to compensate for their lesser mathematical talent by investing effort. The misconception of *girls' non-compensability* puts girls' un-ability to compensate for their lack of talent into focus. Therefore, we expected the MGMQ data to fit a three-factor structure of math-gender misconceptions better than a general-factor structure with one homogeneous misconception construct in a confirmatory factor analysis.

Prevalence Hypothesis

We expect student teachers to rather endorse the first two of the three potential misconceptions. Given the high prominence and face validity of the idea that girls think more empathically whereas boys think more systematically (Empathizing-Systemizing theory; Baron-Cohen, 2005), some student teachers may also believe that these thinking differences are related to worse mathematical abilities – a misconception (*empathizing-systemizing* misconception). Further, we expect some student teachers to endorse the *girls' compensation* misconception referring to the belief that girls only succeed in math because they work hard, whereas boys who succeed are talented. This belief is likely to exist among teachers, because teachers attribute girls' better math grades than boys' math grades to the girls' greater effort (Sáinz et al., 2020). Further, teachers perceive girls only as similarly math-competent as boys if girls work harder (Robinson-Cimpian et al., 2014). Likewise, teachers attribute girls' weak mathematical performance to lacking talent, and boys' weak mathematical performance to lacking effort (Tiedemann, 2002). This research also suggests that (student) teachers may endorse the *girls' non-compensability* misconception to a lesser degree than the *girls' compensation* misconception.

Association Hypothesis

We first expect the three math-gender misconceptions to relate positively with the common math-gender stereotype found in previous research using a simple *female-to-male*-rating for math (for a similar measure, see Nosek, 2007, Nosek et al., 2010). We expect this association, as there are similarities and overlaps amongst math-gender stereotypes and math-gender misconceptions (Klineberg, 1951; Chi and Roscoe, 2002; Kollmayer et al., 2020). More specifically, we expect math-gender stereotypes to be partly based on math-gender misconceptions, which should be expressed in a moderate to high correlation between the two. Secondly, we expect that holding the girls' non-compensability misconception will relate positively with holding fixed-ability mindsets for mathematics (Leslie et al., 2015). Holding the girls' non-compensability misconception means assuming that girls' lack of talent cannot be compensated for, and is thus fixed. This misconception is similar to the idea of fixed ability mindsets for mathematics.

MATERIALS AND METHODS

Participants and Recruiting

A total of 303 student teachers [242 women, 61 men, M_{age} = 21.73 (SD = 4.7, range = 18–51 years)] completed our online survey without dropping out. These data sets were complete (no missing data amongst them). The student teachers had studied on average for 2.28 semesters (SD = 2.28, range = 2–16 semesters). The student teachers' school subjects were mostly German (n = 146) and math (n = 118), followed by other common subjects (e.g., English, biology, politics and economics, philosophy, geography, languages such as French, Spanish, or Latin). More than half of the participants (168; 55.5%) studied at least one STEM subject. Participants were studying to teach at the elementary (n = 79) or secondary school level (n = 191). Some participants were studying to teach in vocational education (n = 7) or special needs education (n = 63). Participants from all over Germany took part in this study; most were from Hessen. The participants, on average, held positive views about gender equality and feminism (M = 3.61, SD = 0.84; scale of 1 = not at all to 5 = very).

The communicated topic of the study was "Mathematics and Gender." The online survey completion was possible between May and July of 2021. We recruited participants via teacher education lectures and seminaries as well as via acquaintances. In total, 360 people clicked on the survey link, of which 303 participants (84.2%) completed the survey. Two people declined consent; the other 55 participants (15.3%) dropped out during the study and were not included in our analyses, yielding the final sample of 303 student teachers.

Study Instruments

Math-Gender Misconception Questionnaire

The self-developed Math-Gender Misconceptions Questionnaire (MGMQ; see **Table 1** for an English translation of the misconception items and **Supplementary Appendix A** for the German original containing all items) served as our main study instrument. It consisted of 30 items. These items comprised statements that participants first must answer with "I disagree" or "I agree" (i.e., verification). Second, each statement comprised a five-point Likert-scale assessing the participants' certainty of having correctly responded to the current statement. The answer options were *very certain*, *certain*, *somewhat certain*, *uncertain*, *very uncertain* (i.e., certainty rating). Certainty ratings were horizontally aligned and presented below the corresponding verification part (see **Figure 1**, for an example item). These two ratings per item are crucial for assessing misconceptions: Holding a misconception should reflect in *incorrect* answers made with a (relatively)

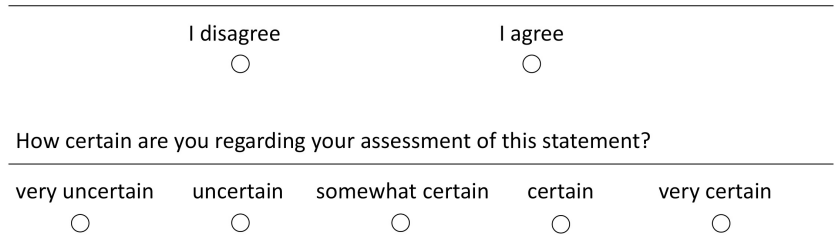


FIGURE 1 | Example item.

high certainty. Incorrect answers with low certainty would rather reflect missing conceptions (see Eitel et al., 2021, for the argumentation). Of the 30 items in the MGMQ, 15 items targeted math-gender misconceptions (see **Table 1**) and 15 items were filler items. Of the 15 misconception items, always five referred to each of the three hypothesized misconceptions (*empathizing-systemizing*, *girls' compensation*, *girls' non-compensability*). The correct answer was to disagree with the misconception items.

The misconception items asked for all of the characterizing aspects of each hypothesized math-gender misconception, by also referring to research findings (Dweck, 1999; Tiedemann, 2002; Muenks et al., 2020; Sáinz et al., 2020), and academic as well as non-academic resources (Baron-Cohen, 2005; Escovar et al., 2016). For the *empathizing-systemizing* misconception, we constructed each of the items to address the combination of the following two stereotypical beliefs: (1) boys are better at math than girls (2) because boys think more systematically, whereas girls think more empathically. We constructed such complex items because only the combination of the two stereotypical beliefs [(1) gender differences in empathizing-systemizing and (2) their direct relation to mathematics performance] is a misconception. An example item was "As girls think rather empathically and boys think rather systematically, boys are on average more talented in mathematics than girls." The same rationale for constructing items applies to the two other misconceptions.

For the *girls' compensation* misconception, items focused both on (1) the belief about gender differences in math talent and on (2) the beliefs that either girls compensate for their fewer talent through hard work or teachers compensate for girls' fewer talent by treating them differently than boys (e.g., more support). An example item regarding *girls' compensation* was "To achieve equally good grades in math, boys have to make less effort because they are more talented than girls are."

For the *girls' non-compensability* misconception, items focused on both (1) the belief about girls being unable to compensate for their lack of talent even with hard work and (2) the belief about implications of this non-compensability in the treatment of genders (such as grading the girls more generously). An example item regarding *girls' non-compensability* was "Despite their on average stronger effort, girls are normally less proficient in math than boys."

We intentionally formulated the misconception items as false statements to gain direct information as to whether the student teachers endorsed this particular misconception. Specifically, disagreeing with a correct statement ("the Earth is a sphere") does not give direct information regarding the underlying misconception (the Earth could be flat, rectangular, a semi-sphere, etc.), whereas agreeing with the incorrect statement ("the Earth is flat") provides direct information about endorsing this particular misconception (cf. Eitel et al., 2021).

The remaining 15 filler-items described true statements related to the math-gender gap, thus they were not misconceptions. An example filler item was "Amongst girls, math is more disliked than amongst boys." The correct answer was to agree with these filler items. The filler-items served to balance the questionnaire. In total, 50% of the statements in the questionnaire were true (i.e., filler-items), while the other half of statements was untrue (i.e., misconception-items). We balanced the questionnaire in order to minimize response biases in the form of acquiescence tendencies (Moosbrugger and Kelava, 2012) because participants might think "some statements must be true" and answer accordingly (cf. Eitel et al., 2021).

Prior to inclusion in the questionnaire, an expert on the math-gender gap and an expert on developing questionnaires revised all items. Additionally, a four-member expert panel (one professor, two postdoctoral researchers, and a Ph.D. student from educational psychology) discussed and refined the questionnaire. Furthermore, we evaluated a prior version of this questionnaire within a pilot study with 246 student teachers. Results of this pilot study suggested that not one unitary construct of math-gender misconceptions but three misconceptions scales might best explain the questionnaire responses, namely the scales of *empathizing-systemizing*, *girls' compensation*, and *girls' non-compensability*. Based on these preliminary findings, we constructed the MGMQ with 15 misconception items, as the former version did not have sufficient misconception items per scale.

Other Instruments

Furthermore, we assessed math-gender stereotypes similar to previous research (Nosek et al., 2010) by asking participants to indicate whether they perceived math as female or male. We used only part of the measure applied by Nosek et al. (2010), who assessed implicit and explicit math-gender stereotypes

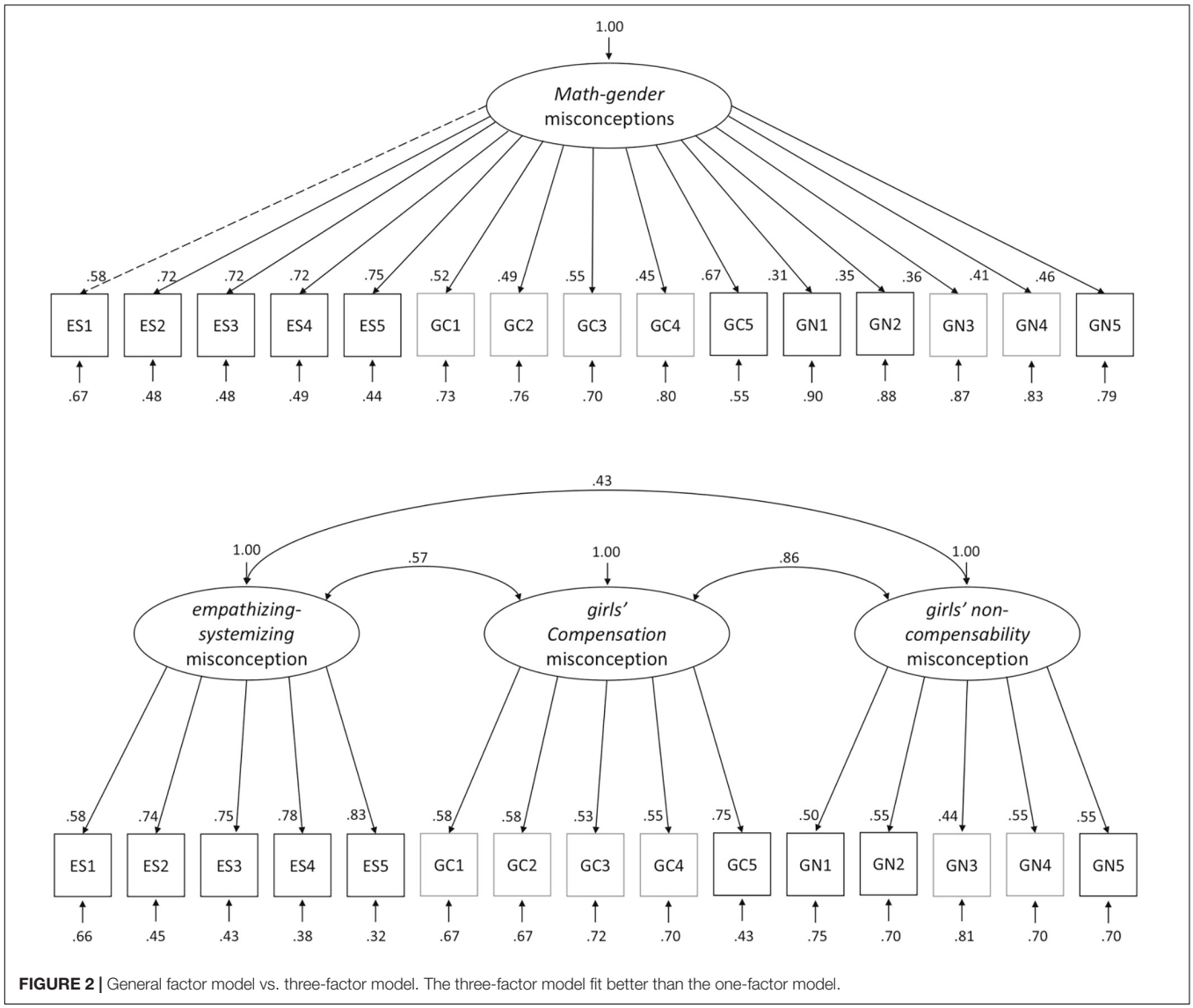


FIGURE 2 | General factor model vs. three-factor model. The three-factor model fit better than the one-factor model.

together with liberal arts-gender stereotypes. Additionally, we extended the scale range to 9 answer options, starting from 1 ("very female") via 5 ("neutral") to 9 ("very male"), to potentially increase variance. The results nevertheless revealed that answers of 1 ("very female"), 8 and 9 ("very male") were outliers in the answer distribution. We thus winsorized the distribution to reduce the biasing effect of the outliers in the correlational analyses.

We then assessed participants' feminism using three items with five-tier Likert-scales each (from *not at all* to *very*). An example item was: "How important is the equality of the genders to you?" The internal consistency of the scale was good ($\omega = 0.79$). For all three items in German and English, see **Supplementary Appendix B**.

We also assessed teachers' *fixed mindset about math ability* with two items adapted from Leslie et al. (2015) and Heyder et al. (2020). An example item was: "Being among the best in math requires a special aptitude that just cannot be taught." Both items were highly correlated ($r = 0.66$, $p < 0.001$) so that we calculated the mean score of both items ($M = 4.07$, $SD = 1.42$).

Before ending the study, participants filled in their demographics such as age, sex, gender, mother language, study subjects, school type they will teach at or already teach at, and semesters studied.

Procedure

When clicking on the web link, participants initially read about the voluntary nature of their participation, that they could end the study whenever they wanted without facing disadvantages, and that we would store all data for 10 years anonymously for the purpose of research only. Participants then gave their informed consent. Participants then read the instruction for the misconception questionnaire, which they then filled in. Then, participants rated how they perceived mathematics on a 9-tier Likert-scale (*female* to *male*). Afterward, participants filled in two items each on fixed mindset in math. They also indicated

their attitude toward feminism. Participants then provided basic demographic information. After participation, we thanked the participants and provided a full debriefing text. Participants took on average 13:44 min (*SD* = 5:48 min) to complete the survey.

Scoring the Misconceptions

We calculated misconception scores by multiplying agreement (coded with +1)/disagreement (coded with −1) and response certainty (coded from 0 = *very uncertain* to 4 = *very certain*; see Eitel et al., 2021). Thereby, we accounted for the nature of misconceptions: Misconceptions are incorrect and are subjectively highly plausible. Thus, if the person assumes an incorrect statement to be more plausible, this person endorses that statement more strongly, reflecting in higher certainty (see Eitel et al., 2021). This stronger endorsement of a misconception is reflected in higher misconception scores (see **Table 1**, for descriptive values). Participants who were very uncertain about an answer (coded with 0), regardless of whether it was correct or not (±1), got a misconception score of 0 (i.e., $\pm 1 \times 0 = 0$), because their (dis-)agreement was probably guessing and indicated no misconception (see Eitel et al., 2021). The stronger participants believed in the misconception, the more certain participants were in their agreement with a false statement (e.g., scores of 2 vs. 4 in the certainty rating). Accordingly, a stronger misconception was indicated by a higher misconception score (e.g., 2 vs. 4). Using this multiplication method, the range of possible values per item was −4 to +4, making it possible to approximate the level of interval-scaled data required to perform confirmatory factor analyses with (robust) maximum likelihood estimation (see Eitel et al., 2021).

We assumed a misconception to be prevalent, whenever participants answered at least one of the five items per misconception scale *incorrectly with high certainty* (i.e., response certainty of 3 or higher, on scale from 0 to 4; see previous section). We did so because a *mixed (mis-)conception* would be prevalent in that case (see Vosniadou, 1994). Misconceptions can be very extreme ("The earth is flat"), but they can also be "alleviated" by integrating correct information ("The earth is round"). However, this alleviation may lead to a so-called *mixed* misconception ("The earth is round, but where we stand on it, it must be flat for us not to fall off"). This would still require further refutation (Vosniadou, 1994). One incorrect answer per misconception scale (made with high certainty) already indicates such a (mixed) misconception, which requires refutation in order to achieve a correct conception (Vosniadou, 1994; see Dersch et al., 2022).

Data Analysis

We used IBM SPSS statistics® for data preprocessing and item statistics. We used R for statistical computing (R Core Team, 2017; version 3.6.23) with the *psych* package for reliability analyses (Revelle and Condon, 2019). We calculated McDonald's omega (ω) for robust reliability estimation even when item-scale correlations are not tau-equivalent (Deng and Chan, 2017). Asymptotic omega simulates the theoretical omega obtained for a test of infinite length with a structure similar to the observed test. Modest reliability for McDonald's omega is at around 0.70 (Nunnally, 1978). However, this convention should be considered with some caution as satisfactory values depend on the measurement purpose (e.g., group statistics or individual assessment) and on the nature of the scale. If assessing broad or heterogeneous constructs, even relatively low coefficients of criterion reliability (e.g., 0.50) do not seriously attenuate validity coefficients (Schmitt, 1996).

We used the *lavaan* package for confirmatory factor analysis (Rosseel, 2012) to inspect the internal structure of the MGMQ by estimating its construct validity. We used maximum likelihood estimation with robust standard errors (MLR) to handle our interval data with moderate deviations from the normal distribution (Li, 2016). We considered the global model fit to be sufficiently good if the following criteria were met: a CFI (comparative fit index) value equal to or higher than 0.95, a root mean square error of approximation (RMSEA) smaller than 0.06 (Hu and Bentler, 1998), and an standardized root mean square residual (SRMR) smaller than or equal to 0.07 (Yu, 2002). We considered the local model fit to be acceptable if values for the fully standardized factor loadings were statistically significant ($p < 0.05$) and higher than 0.30 (Nunnally, 1978; Cristobal et al., 2007).

RESULTS

Structure Hypothesis

We first examined the MGMQ's factorial structure by comparing global and local fit measures of two structural models against each other in a confirmatory factor analysis. We expected the MGMQ data to better fit a correlated three-factor model of math-gender misconceptions (*empathizing-systemizing, girls' compensation,* and *girls' non-compensability*) than a general-factor model with one misconception construct. Accordingly, results revealed an overall acceptable global fit for the three-factor model (with five items per factor), $CFI = 0.94$, $RMSEA = 0.058$, $SRMR = 0.057$, $\chi^2 = 157.75$, $df = 87$, $p < 0.001$. The factors *girls' compensation* and – *non-compensability* were highly positively correlated to each other ($r = 0.86$, $p < 0.001$), and to *empathizing-systemizing* ($r = 0.72$, $p < 0.001$; $r = 0.51$, $p < 0.001$). Results revealed an unacceptable global fit for the general-factor model, $CFI = 0.80$, $RMSEA = 0.10$, $SRMR = 0.08$, $\chi^2 = 296.98$, $df = 90$, $p < 0.001$. Supporting the structure hypothesis, the model fit of the three-factor model was statistically significantly better than the fit of the general-factor model, $\chi^2(3) = 62.50$, $p < 0.001$. On the level of local model fit, factor loadings were all significant (all $ps < 0.01$) and ranged between 0.44 and 0.83 for the three-factor model ($M = 0.62$, $SD = 0.12$; see **Figure 2**). Scale reliabilities [using McDonald's omega (ω)] were good for *empathizing-systemizing* ($\omega = 0.88$), acceptable for *girls' compensation* ($\omega = 0.76$), and acceptable for *girls' non-compensability* ($\omega = 0.72$).

Prevalence Hypothesis

We expected student teachers to rather endorse the first two of the three gender misconceptions about mathematical abilities. As expected, more student teachers believed

that boys are inherently better in mathematics because they think more systematically (*empathizing-systemizing*; 32.0%), and that girls are only as good in mathematics as boys because they work harder (*girls' compensation*; 26.7%) and that girls cannot compensate for their lower mathematical abilities (*girls' non-compensability*; 17.5%). Overall, 14.2% of student teachers endorsed both the *empathizing-systemizing* and the *girls' compensation* misconception, whereas 44.6% of student teachers endorsed at least one of these two misconceptions. In total, 48.2% of student teachers endorsed at least one of the three misconceptions. However, on average, student teachers had negative misconception values in the MGMQ (see **Table 1**). This indicates that the majority of student teachers – correctly – disagreed with the misconception items and did not hold (strong) math-gender misconceptions.

Association Hypothesis

We expected math-gender misconceptions to be positively associated with the prevalence of math-gender stereotypes. We found that 141 out of 303 student teachers indicated math to be more male than female, yielding a prevalence rate of 46.5%. A total of 150 student teachers (49.5%) indicated math to be equally male and female, whereas only 12 student teachers (4.0%) indicated math to be more female than male. Overall, the latent correlation between math-gender misconceptions and holding the math-gender stereotype was moderate, $r = 0.45$, $p < 0.001$. Descriptively, we found that the *empathizing-systemizing*, $r = 0.43$, $p < 0.001$, and the *girls' compensation* misconception, $r = 0.44$, $p < 0.001$, correlated stronger with holding the math-gender stereotype than the *girls' non-compensability* misconception, $r = 0.25$, $p = 0.01$.

Apart from that, we expected holding a fixed ability mindset for mathematics (Dweck, 1999; Leslie et al., 2015) to correlate more positively with the *girls' non-compensability* than with the *girls' compensation* misconception. We found that student teachers with a stronger fixed ability mindset for mathematics believed more strongly in all three misconceptions ($r = 0.28$, $p < 0.001$), however, not to a stronger degree in the *girls' non-compensability* misconception ($r = 0.22$, $p = 0.004$) than in the *girls' compensation* misconception ($r = 0.28$, $p < 0.001$).

DISCUSSION

Math-gender stereotypes held by important socializers like teachers may be contributing to the underrepresentation of girls and women in STEM (for a review, see Gunderson et al., 2012). The goal of this research was to explore the *specific* misconceptions underlying math-gender stereotypes in a student teacher sample. To this end, we first analyzed the structure and prevalence of three potential misconceptions using the newly developed Math Gender Misconceptions Questionnaire (MGMQ). Afterward, we inspected to what degree holding these misconceptions related to holding math-gender stereotypes, and fixed mindsets about math ability.

Structure of Math-Gender Misconceptions Amongst Preservice Teachers

We constructed the MGMQ to uncover a three-factor structure of misconceptions about gender differences in mathematical abilities that we expected to observe based on prior research: *empathizing-systemizing*, *girls' compensation* and *girls' non-compensability*. We obtained evidence for the supposed tree-factor structure via confirmatory factor analysis. The three-factor model fit the data better than the model assuming one general misconception factor (see **Figure 2**). Math-gender misconceptions are thus expressed through three distinct factors. (1) There is the *empathizing-systemizing* misconception assuming that pre-natal testosterone-exposure levels are lower in girls than in boys, which leads to girls thinking less systematically in relation to more empathically. Girls' less systematic thinking – according to this misconception – leads to girls' lower mathematical abilities (Baron-Cohen, 2005). (2) The *girls' compensation* misconception assume that girls are more hardworking than boys, resulting in their equally good performance in math (e.g., equal grades; Tiedemann, 2002; Sáinz et al., 2020). (3) The *girls' non-compensability* misconception assumes that girls are not only less talented in math – for example due to the *empathizing-systemizing* misconception – but furthermore, they lack the means to compensate for their disadvantage, as math talent is fixed (Dweck, 1999; Leslie et al., 2015).

The *empathizing-systemizing* scale showed good reliability; all items correlated substantially with the construct (see **Table 1**). The *girls' compensation* and - *non-compensability* scales showed acceptable reliabilities. The higher reliability of the *empathizing-systemizing* scale, compared to the other two scales, may be due to the items of *empathizing-systemizing* being very homogeneous; they all referred to the explanation of talent differences in boys and girls in mathematics. Items on the other two scales referred to both the talent differences in boys and girls in mathematics and the consequences of such talent differences. Items on the *girls' compensation* scale refer to (1) girls having less talent in mathematics, and (2) girls usually compensating for their lesser talent. Items on the *girls' non-compensability* scale refer to (1) girls having less talent in mathematics, and (2) how girls should be treated to adapt to their lack of talent (lower standards for girls; see **Table 1**, for an overview of all items). Meaning, *girls' compensation* as well as *girls' non-compensability* are broader and more heterogeneous constructs, which may explain their lower reliability coefficients than for the *empathizing-systemizing* scale.

Prevalence and Correlates of Math-Gender Misconceptions

Almost half of the preservice teachers (48.2%) held at least one of the three misconceptions. A majority of student teachers, however, held no math-gender misconceptions, even according to the strict criteria we applied. This finding led to negative

average math-gender misconception scores among student teachers in this sample (see **Table 1**), which imply that on average, math-gender misconceptions are not (strongly) prevalent. These results are encouraging, even if they are still far from ideal. The prevalence of math-gender misconceptions among a subgroup of student teachers is still worrying, since even endorsing just one the misconceptions can affect teachers' instruction. As a consequence, misconceptions may cause a different treatment of the genders (e.g., Carlana, 2019), and reinforce math-gender stereotypes among schoolchildren (e.g., Geis, 1993; Eccles, 2011; Gunderson et al., 2012). The math-gender stereotypes weaken female representation in mathematical careers (e.g., Eccles, 2011; Wang and Degol, 2017). As teachers function as multipliers of their own knowledge and beliefs and teach many students during their career, misconceptions deserve attention and interventions in teacher education and training, even if only a subgroup of teachers seems to endorse such misconceptions.

As expected, both the *empathizing-systemizing* (32.0%) and *girls' compensation* misconception (26.7%) seemed to be more prevalent than the *girls' non-compensability* (17.5%) misconception. This difference in prevalence may partially be due to social desirability. Agreeing to the *empathizing-systemizing* misconception may be more socially desirable than agreeing to statements on the two other misconception scales, because the former statements (1) highlight girls' empathic and social abilities and (2) provide an explanation for girls' lack of talent that did not blame the girls themselves, but rather their genes or pre-natal influences on their body. Like for benevolent sexism (Glick and Fiske, 1996, 1997) these two apparently "positive" beliefs about girls might have been more socially acceptable than agreeing with the beliefs captured by the other two misconceptions.

The other two misconceptions consisted of statements displaying obvious, less benevolent sexism, such as indicating that (1) girls lack talent and (2) the genders should be treated differently and thus unequally. Such attitudes tend to be rejected nowadays among well-educated students in Western societies, like those in our study sample: Accordingly, the students in our sample indicated moderate to high agreement with feminism, which correlated negatively with misconception endorsement ($r = -0.21$, $p = 0.001$). This lower social desirability thus may have reduced agreement rates with the *girls' compensation* and *girls' non-compensability* scale, even though actual beliefs may differ from what participants indicated. Furthermore, the awareness that fixed mindsets in teachers are detrimental to their students (e.g., Canning et al., 2019, 2021; Heyder et al., 2020) seems to be increasing in (teacher) education (Dweck, 2016). Thus, especially the *girls' non-compensability* scale – theoretically a combination of fixed ability mindset ideas and promoting girls' lesser abilities, might be perceived as socially undesirable, which could have contributed to the (relatively speaking), lowest endorsement rates.

Furthermore, the significant correlation between fixed mindset in math and the *girls' non-compensability* misconception as well as the non-significant correlation with the *girls' compensation* misconception supports the construct validity of the MGMQ's constructs: It is only when abilities are perceived as fixed that there is no way to compensate for low abilities. Since fixed mindsets in math have been found to be detrimental only in terms of female students' intrinsic motivation and ability self-concepts (Heyder et al., 2020), these associations further corroborate the importance of the *girls' non-compensability* misconception for female students' engagement in math.

Our findings also support prior research findings of (preservice) teachers holding explicit math-gender stereotypes (e.g., Li, 1999; Tiedemann, 2002; Cimpian et al., 2016; Sáinz et al., 2020). Also in our study, about half (49.5%) of the preservice teachers held explicit math-gender stereotypes. These explicit math-gender stereotypes were associated with math-gender-misconceptions to a moderate degree ($r = 0.45$), tentatively supporting the idea of math-gender-misconceptions underlying math-gender stereotypes. So far, math-gender stereotypes have been assessed either via implicit association testing (e.g., Nosek et al., 2010; Steffens and Jelenec, 2011), or ratings of whether math is more female than male (e.g., Nosek et al., 2010), or via one to three simple items about talent differences (see Hyde et al., 1990; Gunderson et al., 2012). The current assessment of math-gender misconceptions as a construct underlying math-gender stereotypes is a novel approach to understand and potentially refute math-gender gender stereotypes. In the future, assessing math-gender misconceptions in addition to math-gender stereotypes may facilitate the comprehension of math-gender stereotypes and thus our ability to target both – math-gender stereotypes and math-gender misconceptions. With this reasoning, it is important to note that holding the math-gender stereotype correlated most strongly with holding the *empathizing-systemizing* and the *girls' compensation* misconception. Specifically targeting these misconceptions (e.g., by means of refutation text; Tippett, 2010) may thus be a promising means to reduce not just the specific misconception but also math-gender stereotypes to a certain degree. More research applying more measures for math-gender stereotypes and evaluating their association with math-gender misconceptions is necessary to gain more insights into the association between math-gender misconceptions, implicit and explicit math-gender stereotypes, as well as how they manifest in teacher and student teacher behavior. Additionally, applying more measures of explicit math-gender stereotypes in future research to assess the relations between math-gender stereotypes and math-gender misconceptions should help further validate the MGMQ in future research. Hence, this study is the first of a planned series of studies on the relationship between math-gender misconceptions and math-gender stereotypes.

Limitations and Further Research

In this study, we presented the MGMQ, a novel measure assessing misconceptions about gender differences in math abilities. To the best of our knowledge, this is the first study applying the concept of misconceptions (e.g., Eitel et al., 2019) to the important field of women's underrepresentation in math. Therefore, some limitations and questions for future research emerged.

First of all, as the main objective of this research was the construction and evaluation of the MGMQ, we implemented only *one* measure to assess math-gender stereotypes [similarly applied by Nosek et al. (2010)]. It is certainly useful to relate

the MGMQ results to other measures assessing math-gender stereotypes in further research. In this paper, we described the MGMQ development. As the MGMQ has demonstrated its reliability as a measuring tool within our sample, we intend to further research its reliability and interrelations between the MGMQ scales and various implicit and explicit – as well as behavioral stereotype-measures in future research.

Another limitation refers to the risk of triggering socially desirable responses as discussed before. Furthermore, recognizing and reporting socially undesirable stereotypes may require a certain degree of self-awareness among participants (Nosek, 2007). Some may not have thought about their stereotypes because they were unwilling to. But even though reflection is necessary and social desirability may hinder the readiness to self-report stereotypes, direct self-reporting is still known to work best for assessing stereotypes (Axt, 2018). In future research, some items (e.g., "Since girls are on average less mathematically gifted, they should be assessed with different criteria than boys") could be revised to make them more neutral-sounding. Strongly overlapping items could be excluded, forming a short version of the MGMQ (e.g., "As girls think more empathically whereas boys think more systematically, boys are on average more talented for math than girls"). A short version should be economic and especially practical for applying it to in-service teachers, as they have less time to participate in research. In future studies, it would be also promising for researchers to stress that the MGMQ is a knowledge test, not an attitude test, thus hopefully reducing further answer bias due to social desirability. Future research with the MGMQ could also focus on the prevalence of math-gender misconceptions in math teachers, as math teachers, due to their direct influence on girls' math learning, may contribute especially to the upholding of math-gender misconceptions (and math-gender stereotypes). In this regard, we compared the misconception prevalence between student teachers with and without mathematics as teaching subjects here. We observed small and insignificant differences between students with math ($M = -1.64$, $SD = 1.17$) and without math as teaching subject ($M = -1.48$, $SD = 1.40$), $t(301) = 1.04$, $p = 0.30$.

Further, our sample's gender distribution consisting of 79.9% women does not represent the general population. However, this high percentage of women in our student teacher sample resembles the gender distribution of teachers in Germany: The Federal Office for Statistics in Germany assessed teachers' gender in the school year of 2019/2020 and found that 73.1% of teachers in general education were female. As gender might still have influenced the math-gender misconception prevalence, we compared the prevalence rates between genders, and revealed that the prevalence of math-gender misconceptions did not differ between female participants ($M = -1.56$, $SD = 1.33$) and male participants ($M = -1.46$, $SD = 1.27$), $t(301) = 0.54$, $p = 0.58$. This insignificant difference may be due to the exposure to math-gender misconceptions in our society regardless of gender.

Furthermore, implicit and behavioral measures could support the assessment and generate additional knowledge about the prevalence of math-gender misconceptions or math-gender stereotypes.

The goal of the MGMQ is to identify math-gender misconceptions that potentially underlie math-gender stereotypes. As (math-gender) stereotypes have rarely been successfully reduced (FitzGerald et al., 2019; Kollmayer et al., 2020), identifying underlying math-gender misconceptions is a starting point for conceptual change – and hopefully attitude change as well. Interventions targeting misconceptions among teachers (e.g., refutation texts; Menz et al., 2021) could therefore also be applied to revise or reduce stereotypes among teachers.

In addition to the math-gender misconceptions discussed here, there are misconceptions and ideas associated with other stereotypes that influence math representation and warrant research. This should yield insights on whether such associations between stereotypes and misconceptions are specific to the gender topic, or generalizable. One example would be math-race stereotypes (Starr and Simpkins, 2021). The intersectionality of stereotypes, meaning people belonging to more than one minority group (e.g., Black and female) and thus suffering from different overlapping adverse stereotypes, should be considered in future research (Yuval-Davis, 2006; Parker et al., 2020).

Conclusion

This study describes a newly developed instrument assessing misconceptions about gender differences in math ability that potentially underlie gender stereotypes, and which therefore may contribute to the underrepresentation of women in math careers. Our results show that (a) our newly developed questionnaire reliably assessed three distinct misconceptions related to gender differences in mathematics in the first sample, (b) almost half of the participating preservice teachers endorsed at least one of the three misconceptions, whereas a majority did not, and (c) holding these misconceptions was substantially associated with holding math-gender stereotypes.

Identifying the specific misconceptions potentially behind math-gender stereotypes is a good starting point for interventions aiming at conceptual change (Larkin, 2012), also in the field of gender and STEM. Since misconceptions hinder the acquisition of scientifically accurate conceptions (Eitel et al., 2021), overcoming them is important to reduce gender disparities in STEM in the future. This study provides the basis upon which to develop specific instructions in the form of refutation texts during teacher education or training (Eitel et al., 2019; Menz et al., 2021; Dersch et al., 2022).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Lokale Ethik-Kommission des

Fachbereichs 06 der Justus-Liebig-Universität Gießen. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

A-SD, AH, and AE contributed to the conception and design of the study and wrote the sections of the manuscript. A-SD organized the database and wrote the first draft of the manuscript. A-SD and AE performed the statistical analysis. All authors contributed to manuscript revision, read, and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.820254/full#supplementary-material

REFERENCES

Avitzour, E., Choen, A., Joel, D., and Lavy, V. (2020). *On the Origins of Gender-Biased Behavior: The Role of Explicit and Implicit Stereotypes* (No. w27818). Cambridge, MA: National Bureau of Economic Research, doi: 10.3386/w27818

Axt, J. R. (2018). The best way to measure explicit racial attitudes is to ask about them. *Soc. Psychol. Personal. Sci.* 9, 896–906. doi: 10.1177/1948550617728995

Baron-Cohen, S. (2005). The essential difference: the male and female brain. *Phi Kappa Phi Forum* 85, 23–26.

Bian, L., Leslie, S. J., and Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science* 355, 389–391. doi: 10.1126/science.aah6524

Bloom, H. S., Hill, C. J., Black, A. B., and Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *J. Res. Educ. Effect.* 1, 289–328. doi: 10.1080/19345740802400072

Canning, E. A., Muenks, K., Green, D. J., and Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Sci. Adv.* 5:eaau4734. doi: 10.1126/sciadv.aau4734

Canning, E. A., Ozier, E., Williams, H. E., AlRasheed, R., and Murphy, M. C. (2021). Professors Who Signal a Fixed Mindset About Ability Undermine Women's Performance in STEM. *Soc. Psychol. Person. Sci.* 2021:19485506211030398. doi: 10.1177/19485506211030398

Carlana, M. (2019). Implicit Stereotypes: evidence from Teachers' Gender Bias. *Q. J. Econ.* 134, 1163–1224. doi: 10.1093/qje/qjz008

Chi, M. T., and Roscoe, R. D. (2002). The processes and challenges of conceptual change. *Iss. Theory Pract.* 2002, 3–27. doi: 10.1007/0-306-47637-1_1

Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., and Miller, E. K. (2016). . Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *AERA Open* 2:23328588416673617. doi: 10.1177/2332858416673617

Cristobal, E., Flavian, C., and Guinaliu, M. (2007). Perceived e-service quality (PeSQ): measurement validation and effects on consumer satisfaction and web site loyalty. *Manag. Serv. Q.* 17, 317–340. doi: 10.1108/09604520710744326

Deng, L., and Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educ. Psychol. Meas.* 77, 185–203. doi: 10.1177/0013164416658325

Dersch, A., Renkl, A., and Eitel, A. (2022). Personalized refutation texts best stimulate teachers' conceptual change about multimedia learning. *J. Comput. Assist. Learn.* doi: 10.1111/jcal.12671

Dweck, C. (1999). *Self-Theories: Their role in motivation, personality and development*. Hove: Taylor and Francis/Psychology Press.

Dweck, C. (2015). Carol Dweck revisits the growth mindset. *Educ. Week* 35, 20–24. doi: 10.1111/bjep.12072

Dweck, C. (2016). What having a "growth mindset" actually means. *Harv. Bus. Rev.* 13, 213–226.

Eccles, J. (2011). Gendered educational and occupational choices: applying the Eccles et al. model of achievement-related choices. *Internat. J. Behav. Dev.* 35, 195–201. doi: 10.1177/0165025411398185

Eccles, J., Adler, T. F., Futtermann, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). "Expectations, values and academic behaviors," in *Perspecitve on achievement and achievement motivation*, ed. J. T. Spence (Cambridge, MA: Academic Press), 75–146.

Eitel, A., Dersch, A.-S., and Renkl, A. (2019). Wissenschaftliche Annahmen und Fehlkonzepte angehender Lehrkräfte über das Lernen mit Multimedia. *Unterrichtswissenschaft* 47, 451–474. doi: 10.1007/s42010-019-00049-4

Eitel, A., Prinz, A., Kollmer, J., Niessen, L., Russow, J., Ludäscher, M., et al. (2021). The misconceptions about multimedia learning questionnaire. An empirical evaluation study with teachers and student teachers. *Psychol. Learn. Teach.* 20, 420–444. doi: 10.1177/14757257211028723

Eliot, L. (2010). *Pink brain, blue brain: How small differences grow into troublesome gaps-and what we can do about it*. New York, NY: Simon and Schuster, doi: 10.14507/er.v0.1310

Else-Quest, N. M., Hyde, J. S., and Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychol. Bull.* 136, 103–127. doi: 10.1037/a0018053

Escovar, E., Rosenberg-Lee, M., Uddin, L. Q., and Menon, V. (2016). The empathizing-systemizing theory, social abilities, and mathematical achievement in children. *Sci. Rep.* 6, 1–11. doi: 10.1038/srep23011

FitzGerald, C., Martin, A., Berner, D., and Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC Psychol.* 7, 1–12. doi: 10.1186/s40359-019-0299-7

Geis, F. L. (1993). "Self-fulfilling prophecies: A social psychological view of gender," in *The psychology of gender*, eds A. E. Beall and R. J. Sternberg (New York, NY: Guilford Press), 9–54.

Glick, P., and Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *J. Personal. Soc. Psychol.* 70:491. doi: 10.1037/0022-3514.70.3.491

Glick, P., and Fiske, S. T. (1997). Hostile and benevolent sexism: Measuring ambivalent sexist attitudes toward women. *Psychol. Women Q.* 21, 119–135. doi: 10.1111/j.1471-6402.1997.tb00104.x

Gottfredson, L. S. (1998). The general intelligence factor. *Sci. Am.* 9, 24–29.

Greenberg, D. M., Warrier, V., Allison, C., and Baron-Cohen, S. (2018). Testing the Empathizing–Systemizing theory of sex differences and the Extreme Male Brain theory of autism in half a million people. *Proc. Natl. Acad. Sci.* 115, 12152–12157. doi: 10.1073/pnas.1811032115

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., and Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psycholog. Rev.* 109:3. doi: 10.1037/0033-295X.109.1.3

Gunderson, E. A., Hamdan, N., Sorhagen, N. S., and D'Esterre, A. P. (2017). Who needs innate ability to succeed in math and literacy? Academic-domain-specific theories of intelligence about peers versus adults. *Dev. Psychol.* 53, 1188–1205. doi: 10.1037/dev0000282

Gunderson, E. A., Ramirez, G., Levine, S. C., and Beilock, S. L. (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles* 66, 153–166. doi: 10.1007/s11199-011-9996-2

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., and Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychol. Sci. Public Inter.* 8, 1–51. doi: 10.1111/j.1529-1006.2007.00032.x

Hand, S., Rice, L., and Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Soc. Psychol. Educ.* 20, 929–945. doi: 10.1007/s11218-017-9408-8

Hannover, B. (2008). "Vom biologischen zum psychologischen Geschlecht: Die Entwicklung von Geschlechtsunterschieden," in *Lehrbuch Pädagogische Psychologie*, ed. A. Renkl (Edison: Huber), 339–388.

Heyder, A., Steinmayr, R., and Kessels, U. (2019). Do teachers' beliefs about math aptitude and brilliance explain gender differences in children's math ability self-concept? *Front. Educ.* 4, 1–11. doi: 10.3389/feduc.2019.00034

Heyder, A., Weidinger, A. F., Cimpian, A., and Steinmayr, R. (2020). Teachers' belief that math requires innate ability predicts lower intrinsic motivation among low-achieving students. *Learn. Instruct.* 65:101220. doi: 10.1016/j.learninstruc.2019.101220

Holder, K., and Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: a new look from a shifting standards perspective. *Soc. Psychol. Educ.* 20, 471–490. doi: 10.1007/s11218-017-9384-z

Hu, L. T., and Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol. Methods* 3, 424–453. doi: 10.1037/1082-989X.3.4.424

Hughes, S., Lyddy, F., and Lambe, S. (2013). Misconceptions about Psychological Science: a Review. *Psychol. Learn. Teach.* 12, 20–31. doi: 10.2304/plat.2013.12.1.20

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., and Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: a meta-analysis. *Psychol. Wom. Q.* 14, 299–324. doi: 10.1111/j.1471-6402.1990.tb00022.x

Keller, C. (2001). Effect of Teachers' Stereotyping on Students' Stereotyping of Mathematics as a Male Domain. *J. Soc. Psychol.* 141, 165–173. doi: 10.1080/00224540109600544

Kersey, A. J., Csumitta, K. D., and Cantlon, J. F. (2019). Gender similarities in the brain during mathematics development. *NPJ Sci. Learn.* 4, 1–7. doi: 10.1038/s41539-019-0057-x

Klineberg, O. (1951). The scientific study of national stereotypes. *Internat. Soc. Sci. Bull.* 3, 505–514.

Kollmayer, M., Schultes, M. T., Lüftenegger, M., Finsterwald, M., Spiel, C., and Schober, B. (2020). REFLECT-a teacher training program to promote gender equality in schools. *Front. Educ.* 5:136. doi: 10.3389/feduc.2020.00136

Lahance, J. A., and Mazzocoo, M. M. (2006). A longitudinal analysis of sex differences in math and spacial skills in primary school age children. *Learn. Indiv. Diff.* 16, 195–2016. doi: 10.1016/j.lindif.2014.03.020

Larkin, D. (2012). Misconceptions about "misconceptions": Preservice secondary science teachers' views on the value and role of student ideas. *Sci. Educ.* 96, 927–959. doi: 10.1002/sce.21022

Leslie, S. -J., Cimpian, A., Meyer, M., and Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 347, 262–265. doi: 10.1126/science.1261375

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949. doi: 10.3758/s13428-015-0619-7

Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: a review. *Educ. Res.* 41, 63–76. doi: 10.1080/0013188990410106

Menz, C., Spinath, B., Hendriks, F., and Seifried, E. (2021). Reducing educational psychological misconceptions: how effective are standard lectures, refutation lectures, and instruction in information evaluation strategies? *Scholar. Teach. Learn. Psychol.* 2021:269. doi: 10.1037/stl0000269

Merton, R. K., and Merton, R. C. (1968). *Social theory and social structure*. New York, NY: Simon and Schuster.

Miller, D. I., Eagly, A. H., and Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *J. Educ. Psychol.* 107:631. doi: 10.1037/edu0000005

Mischel, W. (1966). "A social-leraning view of sex differences in behavior," in *The development of sex differences*, ed. E. E. Maccoby (Redwood City, CA: Stanford University Press), 57–81.

Moosbrugger, H., and Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. New York, NY: Springer, doi: 10.1007/978-3-642-20072-4_1

Muenks, K., Canning, E. A., LaCosse, J., Green, D. J., Zirkel, S., Garcia, J. A., et al. (2020). Does my professor think my ability can change? Students' perceptions of their STEM professors' mindset beliefs predict their psychological vulnerability, engagement, and performance in class. *J. Exp. Psychol. Gen.* 149, 2119–2144. doi: 10.1037/xge0000763

Muntoni, F., and Retelsdorf, J. (2018). Gender-specific teacher expectations in reading—the role of teachers' gender stereotypes. *Contemp. Educ. Psychol.* 54, 212–220. doi: 10.1016/j.cedpsych.2018.06.012

Nosek, B. A. (2007). Implicit–explicit relations. *Curr. Direct. Psychol. Sci.* 16, 65–69. doi: 10.1111/j.1467-8721.2007.00477.x

Nosek, B. A., and Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *Am. Educ. Res. J.* 48, 1125–1156. doi: 10.3102/0002831211410683

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Banaji, M. R., et al. (2010). Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* 18, 36–88. doi: 10.1080/10463280701489053

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York, NY: McGraw-Hill.

Parker, P. D., Van Zanden, B., Marsh, H. W., Owen, K., Duineveld, J. J., and Noetel, M. (2020). The intersection of gender, social class, and cultural context: a meta-analysis. *Educ. Psychol. Rev.* 32, 197–228. doi: 10.1007/s10648-019-09493-1

Passolunghi, M. C., Rueda Ferreira, T. I., and Tomasetto, C. (2014). Math–gender stereotypes and math-related beliefs in childhood and early adolescence. *Learn. Indiv. Diff.* 34, 70–76. doi: 10.1016/j.lindif.2014.05.005

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reilly, D. (2012). Gender, culture and sex-typed cognitive abilities. *PLoS One* 7:e39904. doi: 10.1371/journal.pone.0039904

Reilly, D., Neumann, D. L., and Andrews, G. (2015). Sex differences in mathematics and science achievement: a meta-analysis of National Assessment of Educational Progress assessments. *J. Educ. Psychol.* 107, 645–662. doi: 10.1037/edu0000012

Retelsdorf, J., Schwartz, K., and Asbrock, F. (2015). "Michael can't read!" – Teachers' gender stereotypes and boys' reading self-concept. *J. Educ. Psychol.* 107, 186–194. doi: 10.1037/a0037107

Revelle, W., and Condon, D. M. (2019). Reliability from α to ω: a tutorial. *Psychol. Assess.* 31:1395. doi: 10.1037/pas0000754

Riegle-Crumb, C., and Humphries, M. (2012). Exploring Bias in Math Teachers' Perceptions of Students' Ability by Gender and Race/Ethnicity. *Gend. Soc.* 26:2. doi: 10.1177/0891243211434614

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., and Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Dev. Psychol.* 50, 1262–1281. doi: 10.1037/a0035073

Rodriguez, S., Regueiro, B., Piñeiro, I., Estévez, I., and Valle, A. (2020). Gender differences in mathematics motivation: differential effects on performance in primary education. *Front. Psychol.* 10:3050. doi: 10.3389/fpsyg.2019.03050

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more.Version 0.5–12 (BETA). *J. Stat. Softw.* 48, 1–36. doi: 10.1002/9781119579038.ch1

Sáinz, M., Fàbregues, S., and Solé, J. (2020). Parent and teacher depictions of gender gaps in secondary student appraisals of their academic competences. *Front. Psychol.* 11:573752. doi: 10.3389/fpsyg.2020.573752

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychol. Assess.* 8, 350–353. doi: 10.1037/1040-3590.8.4.350

Skaalvik, E. M., and Rankin, R. J. (1994). Gender differences in mathematics and verbal achievement, self-perception and motivation. *Br. J. Educ. Psychol.* 64, 419–428. doi: 10.1111/j.2044-8279.1994.tb01113.x

Skaalvik, E. M., Federici, R. A., and Klassen, R. M. (2015). Mathematics achievement and self-efficacy: Relations with motivation for mathematics. *Internat. J. Educ. Res.* 72, 129–136. doi: 10.1016/j.ijer.2015.06.008

Smith, A. E., Jussim, L., and Eccles, J. (1999). Do self-fulfilling prophecies accumulate, dissipate, or remain stable over time? *J. Personal. Soc. Psychol.* 77:548. doi: 10.1037//0022-3514.77.3.548

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *J. Exp. Soc. Psychol.* 35, 4–28. doi: 10.1006/jesp.1998.1373

Starr, C. R., and Simpkins, S. D. (2021). High school students' math and science gender stereotypes: relations with their STEM outcomes and socializers' stereotypes. *Soc. Psychol. Educ.* 24, 273–298. doi: 10.1007/s11218-021-09611-4

Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J. Person. Soc. Psychol.* 69:797. doi: 10.1037/0022-3514.69.5.797

Steffens, M. C., and Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles* 64, 324–335. doi: 10.1007/s11199-010-9924-x

Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educ. Stud. Math.* 50, 49–62. doi: 10.1023/A:1020518104346

Tippett, C. D. (2010). Refutation text in science education: a review of two decades of research. *Internat. J. Sci. Educ.* 8, 951–970. doi: 10.1007/s10763-010-9203-x

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and instruction* 4, 45–69. doi: 10.1016/0959-4752(94)90018-3

Wang, M.-T., and Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): current knowledge, implications for practice, policy, and future directions. *Educ. Psychol. Rev.* 29, 119–140. doi: 10.1007/s10648-015-9355-x

Wentzel, K. R. (2014). "Socialization in school settings," in *Handbook of socialization: Theory and research, Second Edition*, eds J. E. Grusec and P. D. Hastings (New York, NY: The Guilford Press), 251–275.

Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* [Unpublished doctoral dissertation]. Los Angeles, CA: University of California.

Yuval-Davis, N. (2006). Intersectionality and feminist politics. *Eur. J. Women Stud.* 13, 193–209. doi: 10.1177/1350506806065752

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fpsyg.2022.835433

Associations Between Children's Numeracy Competencies, Mothers' and Fathers' Mathematical Beliefs, and Numeracy Activities at Home

Anna Mues, Astrid Wirth, Efsun Birtwistle and Frank Niklas*

Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

Children's numeracy competencies are not only relevant for their academic achievement, but also later in life. The development of early numeracy competencies is influenced by children's learning environment. Here, the home numeracy environment (HNE) and parent's own beliefs about mathematics play an important role for children's numeracy competencies. However, only a few studies explicitly tested these associations separately for mothers and fathers. In our study, we assessed mothers' and fathers' mathematical gender stereotypes, self-efficacy and their beliefs on the importance of mathematical activities at home, and tested their associations with parents' numeracy activities and children's numeracy competencies in a sample of N = 160 children (n = 80 girls) with an average age of M = 59.15 months (SD = 4.05). Both, fathers and mothers regarded boys as being more competent in mathematics than girls. Fathers when compared to mothers reported a greater mathematical self-efficacy. Further, only mothers' self-efficacy was associated with the frequency of numeracy activities with the study child. In contrast, only fathers' beliefs on the importance of mathematics was associated with their numeracy activities which, in turn, predicted children's numeracy competencies. However, the non-invariant constructs and varying results lead to the question whether a revision of existing scales assessing parental beliefs and home numeracy activities is needed to investigate differences of mothers and fathers and their potential associations with children's numeracy outcomes.

Keywords: parental beliefs, home numeracy environment, numeracy competencies, gender stereotypes, self-efficacy, importance of mathematical activities at home

OPEN ACCESS

Edited by:
Daniela Barni,
University of Bergamo, Italy

Reviewed by:
Luciano Romano,
European University of Rome, Italy
Lucrezia Cavagnis,
University of Bergamo, Italy
Robin Peterson,
University of Colorado, United States
Nina Anderson,
University of Colorado, United States,
in collaboration with reviewer RP

****Correspondence:***
Anna Mues
Anna.Mues@psy.lmu.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *14 December 2021*
Accepted: *11 March 2022*
Published: *14 April 2022*

Citation:
Mues A, Wirth A, Birtwistle E and Niklas F (2022) Associations Between Children's Numeracy Competencies, Mothers' and Fathers' Mathematical Beliefs, and Numeracy Activities at Home. Front. Psychol. 13:835433. doi: 10.3389/fpsyg.2022.835433

INTRODUCTION

Children's early competencies and their development are supported by different experiences and aspects in their environment and in everyday life (e.g., Burghardt et al., 2020). In addition to kindergarten attendance (Melhuish et al., 2015) and the home learning environment (LeFevre et al., 2009; Anders et al., 2012; Niklas and Schneider, 2014), parents' beliefs, attitudes and expectations (Sonnenschein et al., 2012; Skwarchuk et al., 2014; del Río et al., 2017) are discussed as important predictors of children's early numeracy development.

Children's early numeracy competencies are essential prerequisites for their later mathematics performance, academic achievement, and school success (Duncan et al., 2007; Jordan et al., 2007;

Niklas and Schneider, 2017). Aspects such as child and family characteristics [e.g., sex or socioeconomic status (SES)] further influence children's cognitive development (Niklas and Schneider, 2014).

The home numeracy environment (HNE) focuses on the early numeracy activities of parents and their children at home (LeFevre et al., 2009). Ecological and sociocultural theories emphasize the importance of the HNE for children's mathematical development (Vygotsky, 1980; Bronfenbrenner and Morris, 2006). Recent studies support this association and reported a positive correlation between home numeracy activities and children's numeracy competencies (LeFevre et al., 2009; Skwarchuk, 2009; Niklas and Schneider, 2014, 2017). However, additional relevant factors such as parental beliefs and expectations toward mathematics, have rarely been considered in recent research (del Río et al., 2017). Further, there are currently only a few studies (Tomasetto et al., 2015; del Río et al., 2019, 2020) that investigated potential differences between mothers and fathers, which may offer interesting insights as fathers have only recently become more involved in their children's lives in many countries (Cabrera et al., 2000; Baker, 2014). The present study investigated children's numeracy competencies and took the following factors into account; (1) three types of parental beliefs toward mathematics: gender stereotypes, self-efficacy and beliefs on the importance of mathematical activities at home, (2) differences in mothers' and fathers' beliefs toward mathematics and home numeracy activities, and (3) children's sex.

Children's Early Numeracy Competencies

Children develop various numeracy competencies even before the start of their formal education (e.g., counting, number line estimation or knowledge of numbers and quantities; Krajewski and Schneider, 2009). Early development of academic competencies is highly predictive for success at school and later in life (Duncan et al., 2007; Jordan et al., 2007; Niklas and Schneider, 2017) as well as for children's further mathematical development (Geary et al., 2009; Krajewski and Schneider, 2009; Jordan et al., 2010). However, these competencies vary greatly between children by the time they start school (Gould, 2012) and are influenced by diverse environmental aspects such as the home learning environment (LeFevre et al., 2009, 2010; Niklas and Schneider, 2017; Susperreguy et al., 2021) and parents' beliefs and expectations toward mathematics (Skwarchuk et al., 2014; del Río et al., 2017). For boys and girls in contrast, often no significant differences concerning numeracy-related and language abilities were found at school entry (Niklas and Schneider, 2012; Kersey et al., 2018). Consequently, girls and boys seem to bring along more or less equal abilities at this age regardless of their sex.

Nguyen et al. (2016) suggested that early numeracy abilities are the strongest predictors of later mathematical achievement. Here, children's advanced counting competencies (i.e., counting forward or backward from a given number or counting with cardinality) have been shown to be more predictive than their basic counting competencies (i.e., number recognition or verbal counting). This finding suggests that it is important to promote advanced counting activities and not to focus on basic counting skills only. However, basic skills still need to be considered as they build the basis for some advanced competencies and critical concepts. Further, children's understanding of quantities and number words and arithmetic abilities were also identified as important predictors for later mathematical competencies (Jordan et al., 2006, 2009; Krajewski and Schneider, 2009). Our study investigated children's numeracy outcomes in the context of additional potentially influencing factors (e.g., HNE and parents' characteristics) to elaborate which aspects play an important role for children's numeracy development before developing later mathematical skills.

Home Numeracy Environment and Parental Involvement in Mathematical Outcomes

The HNE is defined as the interaction between children and their parents concerning numeracy activities within the home environment (e.g., playing dice or counting games and exposure to numerical content; e.g., LeFevre et al., 2009; Skwarchuk et al., 2014). The HNE may be differentiated into formal and informal home activities (e.g., LeFevre et al., 2009). Here, aspects such as using number books and active stimulation of number skills, which require dynamic engagement and the intention of parents to teach mathematics to their children are subsumed as formal HNE. In contrast, informal aspects were described as activities that 'incidentally' support children's numeracy abilities such as naturally occurring activities in the home that induce counting or exposure to numbers (e.g., playing mathematical games and involving the child in measuring ingredients).

Playful learning activities motivate and engage children in learning numbers, counting, and reasoning, and also prepare them to advance their mathematical thinking skills (e.g., problem solving, mental representation of numbers; Cohrssen et al., 2014; Niklas et al., 2016). Parents often provide such learning opportunities and thus motivate their children to learn mathematics (Cohrssen and Niklas, 2019; Gasteiger and Moeller, 2021). For example, Cohrssen and Niklas (2019) reported gains in children's mathematical competencies through playing a math game, underlining the importance of parent-child interactions in shared mathematical activities.

Further, LeFevre et al. (2009) showed that children who are often involved in mathematical activities with their parents at home, are more likely to improve their computational efficacy and accuracy while solving mathematical problems (see also Kleemans et al., 2012). However, positive associations (Niklas and Schneider, 2014; Skwarchuk et al., 2014; Susperreguy et al., 2020) and no significant associations (Skwarchuk, 2009; DeFlorio and Beliakoff, 2015; Missall et al., 2015) were found for the formal and the informal HNE and children's mathematical outcomes with the informal HNE seemingly being the better predictor (LeFevre et al., 2009). Consequently, more research on the specific aspects and mechanisms that may support children's mathematical learning in the context of the HNE are necessary (see also Hornburg et al., 2021).

For instance, research identified additional parental factors which can be linked to the HNE and may impact on children's competencies such as parental beliefs, attitudes and expectations (Sonnenschein et al., 2012; Skwarchuk et al., 2014; del Río et al., 2017). It is also of interest, whether numeracy activities at home and children's numeracy outcomes are influenced by further parental factors such as potential differences between mothers and fathers.

Parental Beliefs Toward Mathematics

Research indicates that parental beliefs, expectations and attitudes predict parental numeracy activities and children's numeracy competency development (Skwarchuk et al., 2014; Missall et al., 2015; del Río et al., 2017; Susperreguy et al., 2020), which aligns with the expectancy-value theory by Eccles. This theory assumes that parental beliefs influence children's achievement motivation, their educational aspirations, and their abilities and provide them with experiences at home and in everyday life which are directed by the beliefs of the parents (Eccles et al., 1983; Jacobs et al., 2005). Recent studies also suggest a direct link between parental beliefs, children's self-concept and their mathematical performance (del Río et al., 2019, 2020), indicating that parents' personal beliefs and thoughts may have a tremendous impact on children's perception of their own abilities and thus also on children's academic outcomes in mathematics.

Parental beliefs toward mathematics can simply be defined as parents' interest in mathematics and their feeling of confidence while performing mathematics (Benz, 2012). Further constructs such as gender stereotypes, self-efficacy and the importance of mathematical activities at home are often subsumed under the umbrella-term *beliefs* (e.g., Sonnenschein et al., 2012, 2016; del Río et al., 2017, 2020). Research shows that parents who tend to have positive beliefs regarding mathematics also engage more frequently in formal numeracy practices such as counting pieces of a pie and teaching how to count (Skwarchuk et al., 2014; Missall et al., 2015; del Río et al., 2017).

Further, parents who engage in formal numeracy activities frequently and who enjoy doing mathematics were reported to have higher expectations for both themselves and their children to perform successfully in numeracy tasks (Blevins-Knabe et al., 2000; Kleemans et al., 2012; Skwarchuk et al., 2014; del Río et al., 2017). However, the associations between parental beliefs and expectations, numeracy-related activities at home and children's numeracy outcomes seem to vary between studies (LeFevre et al., 2009; Skwarchuk, 2009; Sonnenschein et al., 2012; del Río et al., 2017). This finding inspired us to investigate the role that parental beliefs toward mathematics (especially gender stereotypes, self-efficacy and the importance of mathematical activities at home) play for parents' numeracy practices and children's early numeracy competencies.

Gender Stereotypes

Gender stereotypes in certain academic areas such as mathematics are often observed in society and they may differ across countries and cultures (Nosek et al., 2009; Breda et al., 2020; Lewis and Lupyan, 2020). As an important part of the societal structure, parents tend to have stereotypes concerning gender and occupation (Breda et al., 2020; del Río et al., 2020). These beliefs are not necessarily developed intentionally, however, they still may impact on children's own beliefs about mathematics and their actual outcomes (Sonnenschein et al., 2012; del Río et al., 2019, 2020). Gender stereotypes in mathematics can simply be described as favoring one gender over another (e.g., boys can do mathematics better than girls, del Río et al., 2020).

In their systematic review, Gunderson et al. (2012) showed that parents' gender stereotypes and their expectations directly affected children's own beliefs, success and achievement in mathematics. These stereotypes do not only impact on their children's own beliefs and development, they also influence how parents engage with their children while doing mathematical activities. For instance, parents tend to engage with sons more often than with daughters (Jacobs et al., 2005; Nosek et al., 2009; Gunderson et al., 2012; del Río et al., 2017). Moreover, del Río et al. (2017) reported that mothers' engagement in advanced numeracy activities differed depending on the sex of their child; that is, mothers engaged with boys more often than with girls. Consequently, recent research indicates that parental gender stereotypes seem to impact on parent-child interactions as well as on children's achievement and development.

Self-Efficacy

Parents' mathematical self-efficacy also plays an important role for their mathematical beliefs, their own mathematical experiences and achievements, and the mathematical interactions with their children (Missall et al., 2015). Self-efficacy describes the interest in and the ability to achieve certain behaviors successfully (Bandura, 1977). Parental mathematical self-efficacy can thus be defined as parents' belief of being able to solve mathematical problems and their belief of being able to influence children's mathematical learning and their environment in a supportive way (Ardelt and Eccles, 2001).

Peacock-Chambers et al. (2017) showed that high levels of parental self-efficacy were associated with a better quality home learning environment. Further, the frequency of informal mathematical activities and its association with children's numerical understanding was mediated by parents' mathematical self-efficacy and their attitudes toward mathematics. Here, parents' mathematical self-efficacy was also linked indirectly to children's arithmetic skills via informal mathematical activities (Vasilyeva et al., 2018).

Parental Beliefs on the Importance of Mathematical Activities

Parental beliefs about the importance of doing mathematical activities at home are regarded as another important factor that is associated with the HNE and children's engagement in mathematical activities (e.g., Sonnenschein et al., 2012). Sonnenschein et al. (2012) analyzed parental beliefs on the importance of mathematical activities at home and their relation with children's mathematical activities at home. Most of the surveyed parents regarded mathematics at home to be very

important and only 14% of the parents considered it as being not so important.

Parents who reported mathematical activities at home to be important, not only had a more positive attitude toward supporting children's mathematical learning, their children also engaged in mathematical activities at home more often (Sonnenschein et al., 2012).

Although the HNE and parental beliefs are associated with children's numeracy competencies, we still do not know much about potential differences between the beliefs of mothers and fathers, and whether they may be associated differentially to children's numeracy skills.

Differences in Mothers' and Fathers' Mathematical Beliefs and the Numeracy-Related Interactions With Their Child

Previous research on parents' numeracy activities and the interactions with their children has usually relied on data reported by mothers only (Saracho and Spodek, 2008). However, there is some research that took both mothers and fathers into account. For instance, del Río et al. (2017) showed that mothers' advanced numeracy-related interactions were a better predictor for children's numeracy outcomes than fathers' interactions, suggesting that mother-child and father-child interactions may support children's learning in different ways. Further, an indirect effect of mothers' expectations toward mathematics and children's numeracy outcomes was found through the advanced numeracy activities mothers provided at home, whilst no such association was detected for fathers' expectations.

Tomasetto et al. (2015) reported a specific role of mothers' math-gender stereotypes concerning their daughters, but not their sons. del Río et al. (2020) further reported that mothers' and fathers' implicit measures both showed a stronger association for mathematics with males than females. Significantly stronger math-gender stereotypes were found for the explicit measures of mothers compared to fathers. In addition, fathers compared to mothers were more convinced that they are good in mathematics.

The scarcity of research that focusses on the differences between mothers and fathers in their beliefs and numeracy activities at home further underlines the need for such studies that examine the relation of these aspects with children's numeracy outcomes.

The Present Study

In recent years, several studies analyzed the home numeracy activities and parental beliefs (e.g., del Río et al., 2017; Susperreguy et al., 2020). However, only a few studies investigated mothers' and fathers' beliefs toward mathematics and potential differences (e.g., Tomasetto et al., 2015; del Río et al., 2017). In addition, few studies considered the associations between children's sex, children's numerical competencies and the HNE simultaneously (del Río et al., 2017, 2020). Therefore, in the present study, we will try to identify aspects that influence children's numeracy competencies. Here, we analyze three types of parents' mathematical beliefs—namely gender-stereotypes, self-efficacy, and parental beliefs on the importance of mathematical activities at home—and the numeracy practices they conduct with their children at home. One key objective is to investigate potential differences in mothers' and fathers' beliefs toward mathematics and their numeracy activities at home and children's numeracy competencies while considering both parents' and children's sex.

Accordingly, we were interested in answering the following four questions:

(1) Do we find measurement invariance for our constructs, when we ask mothers and fathers the same questions?
(2) Do we find differences between mothers' and fathers' beliefs toward mathematics (i.e., concerning gender stereotypes, self-efficacy, and beliefs on the importance of mathematical activities at home)?
(3) Is there an association between these aspects and the numeracy-related activities at home and children's numeracy outcomes?
(4) Do these associations differ for boys and girls?

To answer these questions, we tested the following five hypotheses:

(1) We expected to measure the same constructs (i.e., beliefs and HNE), when we ask mothers and fathers the same questions (i.e., measurement invariance).
(2) We suggest that mothers and fathers will differ significantly in their mathematical gender stereotypes, self-efficacy and the reported importance of mathematical activities at home. Here, we expected mothers to show lower mathematical self-efficacy than fathers (del Río et al., 2019, 2020).
(3) In addition, we hypothesized that mothers who expect boys to excel in mathematics when compared to girls to have a lower mathematical self-efficacy, whereas fathers with the same stereotypes should have a higher mathematical self-efficacy (del Río et al., 2019, 2020).
(4) We expected parents with a greater mathematical self-efficacy who reported less strong gender stereotypes toward mathematics to engage more often in numeracy related activities and to have children who perform better in the numeracy tasks (del Río et al., 2017).
(5) Finally, we expected that the findings and associations will differ significantly dependent on whether the study child is a boy or a girl.

MATERIALS AND METHODS

Sample

We assessed children's numeracy competencies (NumC) in a sample of N = 310 children (n = 160 girls) with an average age of M = 59.36 months (SD = 3.94) and surveyed both, their mothers and fathers concerning their mathematical self-efficacy (SE), gender stereotypes (GS), beliefs on the importance of mathematical activities at home (IOMA), and

the numeracy activities they provide at home to their children (NA). The data was taken from the first measurement point of the second cohort of the EU-funded, 5-year-longitudinal study "Learning4Kids" project (Niklas et al., 2020a). Trained psychologists, educators and research assistants performed the assessments which included standardized numeracy tests to assess children's numeracy competencies. Further, parents were asked to fill in a written survey assessing SE, GS, IOMA, NA, their family background and children's characteristics. Here, children's sex assigned at birth was reported by their parents in our parental survey.

The majority of our sample spoke German as main language (68.1%). Families, whose first language was not German (27.5%) reported 16 different languages as main language and were provided with surveys in their own language when possible (e.g., Turkish, Polish, English, etc.). Before the beginning of the assessments, families were contacted via mail and received a description of the study and the invitation to contact the project team for participation via e-mail or telephone. In a next step, we called all the families who indicated their interest to participate in our study and explained the study requirements and obtained an informal consent. Some of the participating families were recruited through kindergartens. During the family visit, formal consents were collected. All research activities were approved by the European Research Council Executive Agency and the Ethics committee of the Faculty of Psychology and Educational Sciences at the University of Munich.

As our research focuses on differences between fathers' and mothers' mathematical beliefs and interactions and as the great majority of children in our sample lived together with both parents, only children for whom data from both, mothers and fathers were available, were included in the analyses.[1] Accordingly, about half of the sample (children with complete mother-father dyads) were included in the analytic sample (N = 160, n = 80 girls). Children in this subsample had an average age of M = 59.15 months (SD = 4.05). A potentially biased drop-out between the excluded and all other cases was tested with independent t-tests for our study variables (i.e., numeracy activities, numeracy competencies, beliefs, SES, age, sex). No significant differences between the excluded and all other cases were found (all p's > 0.05, $BF_{10} < 0.07$), except for the beliefs on the IOMA of fathers which were significant ($p < 0.05$), but the Bayes Factor was low ($BF_{10} = 1.37$), indicating that the analytic sample seems to be comparable to the total sample.

Measures

Children's Numeracy Competencies

All participating children were assessed with various numeracy tests. We used the "Marko-Screening—mathematics and concepts of calculation before school entry" (MARKO-S; Ehlert et al., 2020) which includes 21 items concerning numbers, cardinality, ordinal number bars and number division, inclusion and relations (Cronbach's α = 0.78 and Mc Donald's ώ = 0.77). Further, addition and subtraction were tested by an adapted version of the calculation subtest of the "Assessment of basic mathematical competencies in kindergarten" (Krajewski, 2018) with eight items (Cronbach's α = 0.70 and Mc Donald's ώ = 0.69). Various subtests from the "Würzburger preschool test: Assessments of literacy and mathematical (precursor) abilities and linguistic competencies in the last year of kindergarten" (Endlich et al., 2017) were applied to assess competencies such as number sequences forward, number sequences backward, number symbol knowledge and knowledge of numerical representations (Cronbach's α = 0.92 and Mc Donald's ώ = 0.92). All of the subtests consisted of eight items, except number sequences backward with six and knowledge of numerical representations with 10 items. Afterward, scales were built from all items for each subtest. Finally, children's numeracy competencies were measured by a latent variable including all numeracy items (Cronbach's α = 0.93 and Mc Donald's ώ = 0.93).

Parental Surveys

Both parents completed our surveys (see survey questions in **Supplementary Material**). The caregiver who was present during the assessments was asked to fill in the main questionnaire which consisted of questions regarding numeracy practices provided at home, family and child characteristics and additionally included questions about the own beliefs toward mathematics. The other parent, who was or was not present during the assessments, was also asked to fill in a survey which only consisted of questions on home numeracy activities and beliefs toward mathematics. The main questionnaire was offered as a paper and pencil survey at home, whereas the additional parental survey was offered either as a paper and pencil survey version or as an online survey version, to assess as many parental pairs as possible.

Numeracy Activities at Home

Parents were asked about informal numeracy-related activities that they do together at home with their children. The NA were measured as a latent variable and contained six items (adapted from Niklas et al., 2016) with questions about parents' involvement in everyday numeracy activities [e.g., "How often do you involve your child in cooking (e.g., counting, weighing, or measuring ingredients)?] (Mothers' Cronbach's α = 0.65 and Mc Donald's ώ = 0.68; Fathers' Cronbach's α = 0.67 and Mc Donald's ώ = 0.65). Parents rated the items on a 5-point Likert scale (e.g., from several times a week to never). Values of 4–0 were assigned accordingly with higher values indicating more frequent NAs and the values were averaged for both, fathers and mothers.

Parental Beliefs

Parental beliefs toward mathematics were assessed with statements concerning parents' own mathematical SE, their GS toward mathematics and the IOMA at home. Parental SE was measured with 6 items (Mothers' Cronbach's α = 0.75 and Mc Donald's ώ = 0.87; Fathers' Cronbach's α = 0.77 and Mc Donald's ώ = 0.87) and included statements such as "In school, I was good at math" (see also Skwarchuk et al., 2014; Missall et al., 2015; Susperreguy et al., 2020). Parental GS were surveyed with three items (e.g., "Girls need less assistance than boys in mathematics") (Mothers' Cronbach's α. = 87 and Mc Donald's ώ = 0.87; Fathers'

[1]One single case of a mother-mother dyad was present in our sample. Consequently, this case had to be excluded for statistical reasons.

Cronbach's α = 0.88 and Mc Donald's ώ = 0.89), that are based on work by Tatto et al. (2012), Tomasetto et al. (2015), and Blömeke et al. (2017). Parents were also asked to evaluate the importance of their child doing mathematical activities at home with three items (e.g., "It is important to me that my child does mathematical activities at home") (Mothers' Cronbach's α = 0.50 and Mc Donald's ώ = 0.50; Fathers' Cronbach's α = 0.62 and Mc Donald's ώ = 0.62) (Sonnenschein et al., 2012). These items were adapted for our study and values of 0–4 were assigned (from "not at all true" to "completely true").

Statistical Analysis

Data analysis was performed by using IBM SPSS Statistics 28.0 (IBM Corp, 2021), JASP 0.16.0.0 (JASP Team, 2021) and Mplus 8.7 (Muthen and Muthen, 2021). Bayes factors were calculated with JASP (Overstall and King, 2014; Morey and Rouder, 2015; JASP Team, 2021).

The percentage of missing values at the item level of childrens' variables as well as mothers' and fathers' variables was low (max. 8.1%). Children's missing values ranged from 0.6 to 8.1%. Mothers had a range of missing values from 1.2 to 3.1% and for fathers, missing values ranged from 0.6 to 1.2%.

First, a multiple-group confirmatory factor analysis (MGCFA) was conducted to test our analyzed constructs for mothers and fathers. Here, we implemented the diagonally weighted least square estimator (DWLS, WLSMV in MPLUS) as this estimator is recommended to be used for categorical ordinal data and its usage leads to more reliable results when ordered Likert scales are applied (Li, 2016, 2021; Lionetti et al., 2016). Next, we tested measurement invariance to check the comparability of the constructs for mothers and fathers (see **Supplementary Material**).

Descriptive statistics of parents' beliefs toward mathematics and their numeracy activities are shown in **Table 1**. Bayesian paired *t*-tests as well as Bayesian repeated-measurement analyses of variance (ANOVA) were applied to test for potential differences between mothers' and fathers' beliefs, their NA, and to check whether and how parents' mean values may vary for boys and girls. Further, Bayesian independent *t*-tests were conducted to investigate how mothers' and fathers' beliefs differ when having a son or a daughter as study child. Additionally, we tested mothers' and fathers' values in GS and the reported IOMA against an expected mean value by using Bayesian one-sample *t*-tests. Here, parents' answers were reported on a 5-point Likert scale from 0 to 4. The expected mean value of 2 for GS would indicate no perceived differences between boys' and girls' mathematical abilities. These analyses are based on the theoretical assumption that boys and girls at kindergarten age do not differ in regard to their numerical abilities (Niklas and Schneider, 2012; Kersey et al., 2018). Consequently, we assume that parents should not attribute more competence to one sex over the other (boys vs. girls). For IOMA, the expected mean value of 2 would indicate, that parents' regard mathematical activities at home neither as important nor as unimportant. Here, we expected parents would regard the IOMA on average above the mean value of 2 (see Sonnenschein et al., 2012). Finally, a multiple-group structural equation model (MGSEM) was used to analyze the associations of our theoretical model (see **Figure 1**). To evaluate the model fit, several goodness-of-fit indices were considered: the root mean square error of approximation (RMSEA, ≤ 0.06), the comparative fit index (CFI, ≥ 0.95), and the standardized root mean squared residuals (SRMR, ≤ 0.08) (Hu and Bentler, 1999). Although we also report the Chi-Square goodness-of-fit statistic (X^2, $p \geq 0.05$), this measure may be oversensitive to minor model misspecifications and sample size (Chen, 2007). Modification indices aligning with theory were considered to improve the model fit. Here, the highest modification indices one after another were added to the model to examine the changes until a sufficient model fit was achieved (Schumacker and Lomax, 2010). All applied modifications are described in our "Results" section.

RESULTS

Construct Validity and Measurement Invariance

First, we evaluated our measurement models for the study variables using MGCFA. The model fit of the theoretical model was good, except for the SRMR which showed a value slightly

TABLE 1 | Descriptive statistics of parental variables for the total analytic sample and subsamples of boys and girls.

	Total					Boys					Girls				
	N	Min	Max	*M*	*SD*	*N*	Min	Max	*M*	*SD*	*N*	Min	Max	*M*	*SD*
NA m	160	0.67	4.00	2.27	0.76	80	0.67	3.67	2.27	0.75	80	0.67	4.00	2.27	0.78
NA f	160	0.17	3.83	2.18	0.74	80	0.17	3.83	2.09	0.74	80	0.67	3.67	2.27	0.73
GS m	156	0.00	3.33	1.47	0.76	78	0.00	3.00	1.41	0.73	78	0.00	3.33	1.54	0.78
GS f	159	0.00	3.33	1.36	0.87	79	0.00	2.33	1.23	0.84	80	0.00	3.33	1.49	0.89
SE m	159	0.00	4.00	2.83	0.99	80	0.00	4.00	2.84	0.94	79	0.00	4.00	2.81	1.05
SE f	160	0.20	4.00	3.13	0.85	80	0.20	4.00	3.05	0.86	80	0.60	4.00	3.20	0.85
IOMA m	157	1.00	4.00	2.94	0.66	79	1.33	4.00	2.84	0.66	78	1.00	4.00	3.03	0.66
IOMA f	159	0.33	4.00	2.81	0.74	79	0.33	4.00	2.70	0.78	80	0.67	4.00	2.91	0.68

N, sample size; Min, minimum; Max, maximum; M, mean; SD, standard deviation; NA, numeracy activities; GS, gender stereotypes; SE, self-efficacy; IOMA, importance of mathematical activities at home; m, mothers; f, fathers.

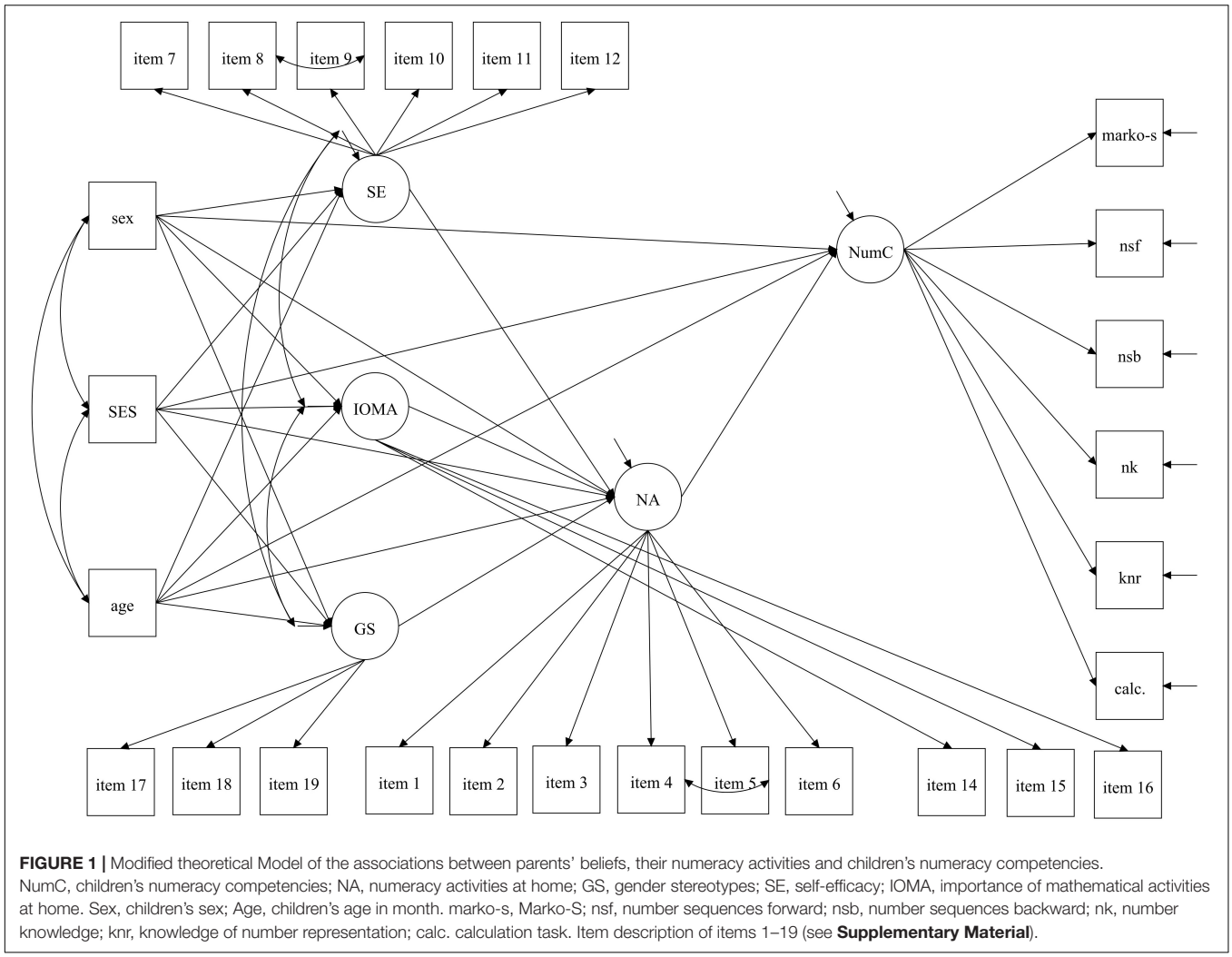


FIGURE 1 | Modified theoretical Model of the associations between parents' beliefs, their numeracy activities and children's numeracy competencies. NumC, children's numeracy competencies; NA, numeracy activities at home; GS, gender stereotypes; SE, self-efficacy; IOMA, importance of mathematical activities at home. Sex, children's sex; Age, children's age in month. marko-s, Marko-S; nsf, number sequences forward; nsb, number sequences backward; nk, number knowledge; knr, knowledge of number representation; calc. calculation task. Item description of items 1–19 (see **Supplementary Material**).

above the cut-off [$X^2(556) = 1024.019$, $p < 0.01$, RMSEA = 0.07, CFI = 0.93, SRMR = 0.08]. To improve the model fit, some modification indices as suggested by MPlus were included after careful theoretical considerations. Here, correlations of item residuals were included by using the WITH statement of MPlus. For mothers' and fathers' numeracy activities, a correlation of the item residuals of item 4 with item 5 was included (see item description in **Supplementary Material**). Further, for parent's self-efficacy, correlations of the item residuals of items 9 and 10 were included (see item description in **Supplementary Material**). With the application of these modification indices, our model showed a slightly better model fit [$X^2(552) = 929.284$, $p < 0.01$, RMSEA = 0.07, CFI = 0.95, SRMR = 0.08]. The Chi-square differentiation test showed a significant p-value ($X^2 = 124.007$, $df = 4$, $p < 0.001$), indicating that we can proceed with this modified model.

In order to evaluate whether we assessed equal constructs for mothers and fathers with our parental survey, measurement invariance was tested. For later analyses (i.e., paired t-tests, and MGSEM), scalar invariance was needed to compare the latent means of mothers and fathers. To evaluate the model fit of the observed data, the change of the alternative Comparative Fit Index (CFI; $\leq$ –0.01) and root mean square error of approximation (RMSEA; $\leq$ 0.015) was used instead of the very sensitive Chi-Square (X^2) (Chen, 2007). For our measurement model, configural invariance was found only. This finding indicates that the results of mothers and fathers cannot be compared in regard to mean difference tests, but that our survey questions seem to measure the same factor structure of our constructs for mothers and fathers (see **Supplementary Material**). We will continue with the planned comparisons between mothers and fathers, but will discuss this limitation later.

Mothers' and Fathers' Beliefs and Home Numeracy Activities

Paired t-tests showed that no significant differences between mothers' and fathers' beliefs and numeracy activities were found, with the exception of SE. Here, fathers showed a significantly greater SE toward mathematics than mothers [fathers: $M = 3.13$, $SD = 0.85$; mothers: $M = 2.83$, $SD = 0.99$; $t(158) = -3.08$; $p < 0.01$; $BF_{10} = 8.14$, Cohen's $d = -0.24$, small effect size].

Further, no significant differences for mothers' and fathers' beliefs and numeracy activities were found when the study child's sex was included in our repeated-measurement ANOVAs. Consequently, no differences for mothers and fathers were found, independent of the sex of the study child.

Further, parents were asked, whether girls have better mathematical competencies and need less support than boys. On average, both mothers and fathers regarded boys to be more competent in mathematics than girls [comparison to expected mean: mothers: $t(155) = -8.68$, $p < 0.01$, Log(BF_{10}) = 27.83, Cohen's $d = -0.70$, medium effect size; fathers: $t(158) = -9.261$, $p < 0.01$, Log(BF_{10}) = 31.34, Cohen's $d = -0.73$, medium effect size]. In addition, parents regarded mathematical activities at home to be important on average [comparison to the expected mean: mothers: $t(156) = 17{,}679$, $p < 0.01$, Log(BF_{10}) = 82.33, Cohen's $d = 1.1$, large effect size; fathers: $t(158) = 13.825$, $p < 0.01$, Log(BF_{10}) = 59.40, Cohen's $d = 1.41$, large effect size].

Associations Between Parents' Beliefs Toward Mathematics, Their Numeracy Activities and Children's Numeracy Competencies

To evaluate the associations between our study variables, a MGSEM was conducted (see **Figure 2**).

Path Analysis

In our model, we expected a direct association between parents' NA and NumC and an indirect effect of mothers' and fathers' beliefs toward mathematics on children's NumC via the NA. Further, we controlled for children's sex, age and families' SES. We used a MGSEM to compare mothers and fathers (see **Figure 1**). Here, the goodness-of-fit indices suggested a good model fit, with the exception of the SRMR [$X^2(676) = 864.049$, $p < 0.001$, RMSEA = 0.04, CFI = 0.97, SRMR = 0.11].

Results from our MGSEM demonstrated that mothers' NA were not significantly ($\beta = 0.096$, $p > 0.05$) associated with children's NumC (see **Figure 2**). The indirect paths of mothers' beliefs on children's NumC via NA were also not significant. However, mothers' reported SE was significantly positively associated with their NA ($\beta = 0.319$, $p < 0.001$). In addition, mothers' IOMA and GS ($\beta = 0.616$, $p < 0.001$) and IOMA and SE ($\beta = 0.251$, $p < 0.01$) were associated, indicating that the attitude of the importance to do mathematical activities at home is accompanied by a more positive attitude that girls are more competent in mathematics than boys. No significant association was found between their GS and their SE. Additionally, children's age was significantly associated with children's NumC ($\beta = 0.367$, $p < 0.001$). Here, older children had better outcomes in comparison to younger children. No significant associations were found for children's sex and families SES and mothers' beliefs and NA.

For fathers, we found a significant association between their NA and children's NumC ($\beta = 0.251$, $p < 0.01$). Contrary to mothers, fathers' beliefs on the IOMA were significantly positively associated with their NA ($\beta = 0.611$, $p < 0.01$). Here, additionally a total indirect effect on children's NumC was found ($\beta = 0.153$, $p < 0.05$), revealing that fathers who value the importance of mathematical activities to a greater extent also engage more often in numeracy activities at home with their children, who—in turn—show better numeracy abilities. No further indirect effects were found for fathers' beliefs, their NA

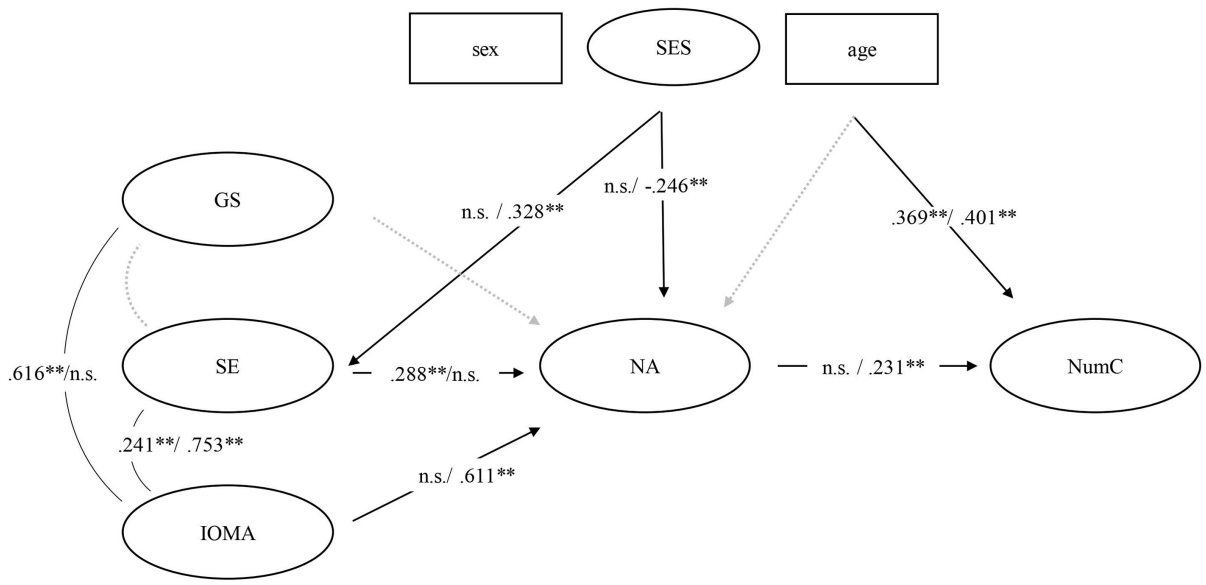


FIGURE 2 | Path analysis of the associations between mothers' and fathers' beliefs toward mathematics, their numeracy activities at home, and numeracy competencies of the children. NumC, children's numeracy competencies; NA, numeracy activities at home; GS, gender stereotypes; SE, self-efficacy; IOMA, importance of mathematical activities at home. Sex, children's sex; Age, children's age in month. Black lines indicate significant associations, and dotted gray lines indicate non-significant associations. $^{**}p < 0.01$, n.s., non-significant. Path coefficients for mothers/fathers.

and NumC. Their SE and GS were not significantly associated with each other and with the NA. However, there was a significant correlation between fathers' SE and their beliefs on the IOMA ($\beta = 0.774$, $p < 0.001$), whereas no such association was found for GS and IOMA. In regard to our control variables, SES was positively associated with fathers' SE ($\beta = 0.338$, $p < 0.001$) and negatively associated with their NA ($\beta = -0.243$, $p < 0.05$). Again, children's age was associated with children's NumC ($\beta = 0.401$, $p < 0.001$). No significant associations were found for children's sex and the other study variables.

DISCUSSION

Children's early numeracy competency development and later mathematical achievement have been investigated previously, however, many questions about influencing factors such as the HNE or parental beliefs remained unanswered (Niklas and Schneider, 2014, 2017; del Río et al., 2017; Susperreguy et al., 2021). This study investigated potential differences between mothers' and fathers' beliefs and numeracy practices at home. Contrary to recent research (Niklas and Schneider, 2014, 2017; Skwarchuk et al., 2014; del Río et al., 2017; Susperreguy et al., 2020), our results only confirmed a significant association of fathers' numeracy practices at home and children's numeracy competencies, but not of mothers' numeracy practices and children's numeracy competencies. Moreover, our results expand current research (del Río et al., 2020; Susperreguy et al., 2020) by contributing to the identification of further factors that may influence children's numeracy competency acquisition and parental practices by examining the distinct associations of mothers' and fathers' beliefs and child outcomes.

The analyses of measurement invariance (H1) showed that the questions we asked mothers and fathers may have assessed the same factor structure of our constructs for mothers and fathers, however, the means were not equivalent. Here, it may be possible that mothers and fathers have a different understanding of our constructs, as found by Meunier and Roskam (2009) who analyzed self-report self-efficacy and found differing associations for mothers and fathers. Consequently, more research is needed to develop standardized parental surveys that provide not only objective, reliable and valid assessments of parental beliefs, but that are also comparable across main caregivers. Such surveys should include questions about the formal and the informal HNE and may also refer to a broader construct of the home math environment by including further aspects of mathematics as proposed by recent literature (e.g., geometry, spatial activities, patterning, and measurement; Zippert and Rittle-Johnson, 2020; Hornburg et al., 2021). As no scalar measurement invariance was found, our reported findings must be interpreted with caution: No direct comparisons of the values of mothers and fathers are possible. Nevertheless, in the following we try to identify possible explanations for the potential differences we found between mothers and fathers.

Our results showed significant differences between mothers' and fathers' beliefs for parents' self-efficacy (H2). As expected, fathers' reported mathematical self-efficacy was greater than that of mothers (Hofmann et al., 2005; del Río et al., 2019, 2020). Here, societal and cultural stereotypes might have differential effects on males and females, as, for example, even if women work in STEM-related occupations, they still show lower self-concept scores in mathematics compared to men (Niepel et al., 2019; Breda et al., 2020; Lewis and Lupyan, 2020).

Another possible explanation could be that mothers seem to have a higher level of math anxiety than fathers, which is reflected in their self-assessment. For example, Schmader et al. (2004) showed that stereotypes can cause anxiety, which in turn has a negative effect on the awareness of one's own abilities. Further, del Río et al. (2017) reported statistically significant higher math anxiety levels for mothers than for fathers, which led to less frequent engagement in mathematical activities. According to Hornburg et al. (2021), cultural influences need to be considered and more research is needed to identify mechanisms that lead to the on average lower SE of mothers. Moreover, as parental attitudes and believes may impact parents' interactions with their children and children's own attitudes (e.g., Niklas et al., 2020b), it is important to inform parents, especially mothers, and to intervene early so that any potential detrimental effects for children can be prevented.

Consistent with prior research, fathers and mothers in our sample regarded boys to be more competent in mathematics than girls (del Río et al., 2019, 2020). These results fit with the literature on mathematical gender stereotypes of adults (Nosek et al., 2009; Miller et al., 2015; Breda et al., 2020) and underline the suggestion of del Río et al. (2017), that cultural stereotypes may implicitly influence parents' own experience of raising their own son or daughter.

As stated before and contrary to current research (Niklas and Schneider, 2014, 2017; Skwarchuk et al., 2014), our findings supported the assumption of a direct association of children's numeracy competencies and parents' numeracy-related activities for fathers only, but not for mothers. This finding stands in contrast to results of del Río et al. (2017), who reported a significant association for mothers only. Here, it should be noted that we only measured informal aspects of the HNE (see LeFevre et al., 2009) and that the associations reported by del Río et al. (2017) were found in the context of formal numeracy practices.

Focusing on the formal HNE or using a more comprehensive assessment of the home math environment as suggested in recent literature (e.g., Zippert and Rittle-Johnson, 2020; Hornburg et al., 2021) may lead to more informative and possibly different results. Such an approach may also help to shed light on the inconsistent findings concerning the association of formal and informal numeracy practices with children's numeracy development (see Elliott and Bachman, 2018). Indeed, analyses with similar measures, but with data from only one main caregiver (mostly mothers) for whom more information about the HNE were assessed, showed that numeracy practices provided at home were a significant predictor of children's numeracy competencies (Mues et al., 2021).

In regard to further associations of parental beliefs, numeracy activities and children's numeracy competencies (H3 and H4), our findings showed that mothers with higher SE provided more

frequent numeracy activities at home for their children. Similar results have been reported by Peacock-Chambers et al. (2017), who demonstrated that higher parental self-efficacy levels were associated with higher scores in measures of the home learning environment. This would also inversely be in line with findings of del Río et al. (2019), who reported that mothers with lower self-efficacy values provided lower quality numeracy activities for their children. Further, Vasilyeva et al. (2018) reported an indirect effect of parents' self-efficacy to children's arithmetic skills via parents' informal mathematical activities.

Additionally, mothers' reported IOMA and SE were significantly correlated. This result indicates that higher SE values and the reported IOMA may be mutually dependent (see also Sonnenschein et al., 2012). Our findings in regard to parental beliefs on the importance of doing mathematical activities at home are in line with findings from Sonnenschein et al. (2012, 2016) and demonstrate that both, mothers and fathers endorse the importance of mathematical activities at home.

In contrast to mothers, a significant association between the IOMA and fathers' NA was found, indicating that the reported belief about the importance of numeracy activities at home is associated with the direct implementation of such activities. We further found an indirect effect of IOMA via NA on children's numeracy abilities. This result aligns with findings from Sonnenschein et al. (2012), who found comparable results, when measuring parental beliefs and the frequency of children's numeracy related activities. The higher SE of fathers compared to mothers may play a role for the different associations found for both main caregivers. Although no direct association of fathers' SE with fathers' NA were found, SE correlated significantly with fathers' reported IOMA. Here, further research is needed to understand the causal associations of our measures and why specific associations were found for mothers or fathers only.

Children's sex, did not show any significant associations with parental beliefs, NA, and child outcomes in any of our analyses (H5; see also De Keyser et al., 2020; Zippert and Rittle-Johnson, 2020). In contrast, for families' SES, significant associations with fathers' SE and NA were found, indicating higher SES is associated with a higher SE of fathers, but lower frequencies of NA. Tazouti and Jarlégan (2019) reported similar findings about the positive association between parents' SES and self-efficacy, but stated that this association was stronger for mothers, contrary to our findings. However, in their analyses, they did not find measurement invariance, and therefore their results must be interpreted with caution.

As fathers are still considered as the main earner in families in many cultural contexts, societal expectations may influence their SE. For instance, gender stereotpyes often lead to women being regarded as less competent compared to men, in particular in the field of science, technology, engineering and mathematics (Niepel et al., 2019; Breda et al., 2020). Finally, our findings are in line with previous research showing a significant association between children's age and their competencies, with older children outperforming younger ones (e.g., Niklas and Schneider, 2017).

However, in addition to parents' beliefs and numeracy activities, there are other contributing factors which have a potential impact on children's numeracy abilities. For instance, Puglisi et al. (2017) did not find a direct association of the informal home literacy environment with children's literacy skills while controlling for parental factors. They argued that parents' genetics may also be an important factor for passing on good numerical and mathematical skills onto children. A genetic component on children's mathematical abilities was also mentioned by Hart et al. (2009), who analyzed data on 314 same-sex twins.

Genetic influences on children's academic achievement are discussed in numerous studies (see e.g., Ludwig et al., 2013; Baron-Cohen et al., 2014; Davis et al., 2014; Pettigrew et al., 2015). Moreover, prior research also discussed intergenerational transmissions between parents and their children concerning mathematical abilities. Here for example, parents' approximation number system was associated with toddlers' number processing, even after controlling for children's vocabulary and parents' mathematical abilities (Navarro et al., 2018). Consequently, it is recommended to consider a genetically sensitive design when investigating children's mathematical abilities in the context of the HNE (Napoli and Purpura, 2018; Hart et al., 2021) and a more differentiated look at parental factors and other influencing aspects is needed. Still, our findings contribute to the understanding of parental aspects that are associated with numeracy-related activities at home and children's numeracy outcomes (**Figure 2**).

The fact that our findings are supported by the results and suggestions of previous research, but are also in contrast with some other research (e.g., Sonnenschein et al., 2012; del Río et al., 2017, 2019; Peacock-Chambers et al., 2017; Vasilyeva et al., 2018), underlines the importance of further investigation of parental factors, such as beliefs and home numeracy activities, but also of differences between mothers and fathers. In our view, surveys and questions assessing parents own beliefs and numeracy practices need to be discussed in the context of missing measurement invariance and the prevailing criticism of using self-reported data only in most studies (Missall et al., 2016; Zippert and Rittle-Johnson, 2020). Our findings implicate that we need to improve our measurement methods before investigating the potential differences between mothers and fathers. Further, we suggest that future research should not only take potential differences between mothers and fathers into account, but should also analyze how parents influence each other in their beliefs and activities and thus may together impact on the development of their children.

Further, clearer definitions of different aspects of parents' beliefs are needed, as we noticed different wordings and definitions for similar items and scales in research. For example, del Río et al. (2020) used the term "parents' self-concept" when using very similar items and questions that we defined as parental self-efficacy. Additionally, other relevant family characteristics (e.g., children's and parent's sex, age, SES etc.) need to be considered and examined concerning their associations with each other, as well as with different aspects of the HNE and children's numeracy competencies (see Hornburg et al., 2021). We would like to stress the importance of developing novel surveys or of improving existing surveys to measure parental mathematical beliefs and the home mathematical environment and potential differences between parents.

Despite these open questions, our findings underline the need for more practical implications to support children's early numeracy development. Here, interventions designed to further investigate gender-based differences, self-efficacy beliefs and HNE may improve our understanding and uncover mechanisms that are at work (see Kaya and Lundeen, 2010). For instance, during a parent evening and through playful parent-child activities at the kindergarten, parents can be made aware of the importance of the role their beliefs and actions play for their children's competency development (Niklas et al., 2016). Another option would be to apply digital interventions, for example via a mobile app which regularly provides useful information or practical tips for parents that they may use in their everyday life together with their children (Niklas et al., 2020a).

Further, a long-term intervention program with early childhood teachers showed change and modifications in beliefs toward mathematics as well as in their pedagogical content knowledge (Bruns et al., 2017). Kaya and Lundeen (2010) reported an effective intervention in the context of parents' attitudes and interest in science by applying an interactive home, school and community collaboration. Their findings showed that family interactions and parents' attitudes toward science became more positive and the interest in the involvement of elementary science increased due to the intervention. Consequently, we assume that interventions including information on the HNE and gender-based differences, and enabling parents to inform themselves about these topics and share their thoughts, feelings and ideas in a group accompanied by a professional will lead to a better understanding and a potential change of their belief set. Such interventions may also influence parental abilities, attitudes, and feelings about mathematics, but more importantly might also improve the relationship with their children and positively impact on children's beliefs and competencies.

Limitations and Further Research

Our study has several limitations that need to be considered when interpreting these findings. First, we only used cross-sectional data, so that no causal interpretation of our findings is possible. Longitudinal data collection and using a mixed-method approach would allow a greater insight onto this topic. However, many of our results align with findings of recent research (Sonnenschein et al., 2012; del Río et al., 2017, 2019, 2020).

Second, due to missing data, we only were able to analyze data from a reduced sample ($N = 160$). There were minor differences between fathers who remained and fathers who dropped out in regard to the IOMA, which needs to be taken into account when interpreting the results.

Third, we could not establish scalar measurement invariance between mothers and fathers as stated in our discussion. Therefore, all comparisons between mothers and fathers need to be interpreted cautiously. However, our descriptive analyses still provide very important information and indicate that studies should consider the sex of both, parents and children. Here, new measurement instruments are needed that work similarly for mothers and fathers (see also Hornburg et al., 2021). In addition, we did not test for potential differences between the different language versions of our survey items and this needs to be considered when interpreting the results of families with a language background other than German.

Fourth, our IOMA scale showed a low internal inconsistency, which might be driven partly by the fact that it only included three items. Here, a more comprehensive assessment of IOMA would be helpful for a more reliable measure.

Fifth, we did not control for siblings of the study child. Consequently, the results for parents need to be interpreted with caution, as siblings of the same or different sex may also influence parental mathematical beliefs and activities as well as children's numeracy competencies.

Sixth, our findings rely on self-reported data of parents only, which may lead to biased and social desirable answers (Missall et al., 2016; Zippert and Rittle-Johnson, 2020). Here, parent-child interactions captured by observational measures or qualitative data on more specific aspects and actions at home assessed through interviews may be useful additional methods of data collection.

In our research, we focussed on mothers and fathers as these were the most common main caregivers for the children in our sample. Here, we only assessed heteronormative families as our total sample included only one case that differed from the majority. However, it would also be of great interest to investigate whether the results would change for same-sex parents or other caregivers (e.g., grandparents).

Seventh, it should be mentioned that the current study did not use a genetically sensitive design. Future research should consider both environmental and genetic factors when investigating associations of children's numeracy abilities and family characteristics.

Finally, it has to be mentioned that the sex of the children was reported as a binary construct in our study only which is consistent with the historical approach in this field but is questioned in latest research regarding its adequacy (see e.g., Berner et al., 2020). However, for the age group analyzed in our study, we still believe that a binary classification will be appropriate for almost all children.

CONCLUSION

Our findings indicate that parents regard boys to be more competent in mathematics than girls. Additionally, parents' self-efficacy differed with mothers showing a lower mathematical self-efficacy compared to fathers. Further, mothers' mathematical self-efficacy and fathers' reported importance of mathematical activities at home correlated with actual numeracy activities at home. Only the frequency with which fathers engaged in numeracy activities with their child were positively associated with children's numeracy competencies.

Moreover our findings raise very important questions for the field of educational psychology: What do we measure when we assess mathematical beliefs and activities of one main caregiver via survey only? How would results differ when both main caregivers are surveyed with questionnaires that show scalar

measurement invariance? Are the findings valid for both, boys and girls, or do we need to put a greater focus on parental and child sex differences from early age onward?

Our results indicate that we are still in need of better, standardized and thoroughly evaluated assessment tools (see also Hornburg et al., 2021). Further, more research on the various influencing factors and their interaction in the context of children's numeracy competency development is needed. Our findings demonstrate that there may be relevant differences between mothers' and fathers' beliefs and numeracy activities at home, which need to be considered for a better understanding of children's early numeracy development.

The main goal should be to support children's competencies development regardless of the main caregiver's and the child's sex. Consequently, we also need more detailed information about existing differences and about how best to support children and their parents according to the individual needs of the child. Future research should consider and analyze practical implications which will provide more insight into topics such as HNE and beliefs toward mathematics and which will lead to parental awareness on the importance of the role their beliefs and actions play for their children's competencies development.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the European Research Council Executive Agency and Ethics Committee of the Faculty of Psychology and Educational Sciences at the University of Munich. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AM and FN conceptualized the main ideas of the manuscript. AM conducted the main analyses and wrote the original draft. AM, AW, EB, and FN investigated the study, reviewed, and edited the manuscript. FN was responsible for the resources, supervision, project administration, and funding. EB was responsible for the data curation. All authors have read and agreed to the published version of the manuscript.

FUNDING

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (Grant agreement no. 801980). The design of the study, collection, analysis and interpretation of data and writing the manuscript were not influenced by the ERC.

ACKNOWLEDGMENTS

We are grateful to all the children and their parents who participated in this study and to our research assistants who contributed to our research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.835433/full#supplementary-material

REFERENCES

Anders, Y., Rossbach, H.-G., Weinert, S., Ebert, S., Kuger, S., Lehrl, S., et al. (2012). Home and preschool learning environments and their relations to the development of early numeracy skills. *Early Child. Res. Q.* 27, 231–244. doi: 10.1016/j.ecresq.2011.08.003

Ardelt, M., and Eccles, J. S. (2001). Effects of mothers' parental efficacy beliefs and promotive parenting strategies on inner-city youth. *J. Fam. Issues* 22, 944–972. doi: 10.1177/019251301022008001

Baker, C. E. (2014). African American fathers' contributions to children's early academic achievement: evidence from two-parent families from the early childhood longitudinal study–birth cohort. *Early Educ. Dev.* 25, 19–35. doi: 10.1080/10409289.2013.764225

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.*84, 191–215. doi: 10.1037/0033-295x.84.2.191

Baron-Cohen, S., Murphy, L., Chakrabarti, B., Craig, I., Mallya, U., Lakatošová, S., et al. (2014). A genome wide association study of mathematical ability reveals an association at chromosome 3q29, a locus associated with autism and learning difficulties: a preliminary study. *PLoS One* 9:e96374. doi: 10.1371/journal.pone.0096374

Benz, C. (2012). Attitudes of kindergarten educators about math. *J. Für Mathematik Didaktik* 33, 203–232. doi: 10.1007/s13138-012-0037-7

Berner, N., Rosenkranz, L., and Schütz, J. (2020). "Geschlecht als variable in der quantitativen bildungsforschung [Gender as a variable in educational research]," in *Methoden und Methodologien in der Erziehungswissenschaftlichen Geschlechterforschung*, eds M. Kubandt, M. Kubandt, and J. Schütz (Opladen: Verlag Barbara Budrich), 26–44.

Blevins-Knabe, B., Austin, A. B., Musun, L., Eddy, A., and Jones, R. M. (2000). Family home care providers' and parents' beliefs and practices concerning mathematics with young children. *Early Child Dev. Care* 165, 41–58. doi: 10.1080/0300443001650104

Blömeke, S., Dunekacke, S., and Jenßen, L. (2017). Cognitive, educational and psychological determinants of prospective preschool teachers' beliefs. *Eur. Early Child. Educ. Res. J.* 25, 885–903. doi: 10.1080/1350293X.2017.1380885

Breda, T., Jouini, E., Napp, C., and Thebault, G. (2020). Gender stereotypes can explain the gender-equality paradox. *Proc. Natl. Acad. Sci. U. S. A.* 117, 31063–31069. doi: 10.1073/pnas.2008704117

Bronfenbrenner, U., and Morris, P. A. (2006). "The bioecological model of human development," in *Handbook of Child Psychology: Theoretical Models of Human Development*, eds W. Damon and R. M. Lerner (Hoboken, NJ: Wiley).

Bruns, J., Eichen, L., and Gasteiger, H. (2017). Mathematics-related competence of early childhood teachers visiting a continuous professional development course: an intervention study. *Math. Teach. Educ. Dev.* 19, 76–93.

Burghardt, L., Linberg, A., Lehrl, S., and Konrad-Ristau, K. (2020). The relevance of the early years home and institutional learning environments for early mathematical competencies. *J. Educ. Res. Online* 12, 103–125.

Cabrera, N. J., Tamis-LeMonda, C. S., Bradley, R. H., Hofferth, S., and Lamb, M. E. (2000). Fatherhood in the twenty-first century. *Child Dev.* 71, 127–136. doi: 10.1111/1467-8624.00126

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equat. Model.* 14, 464–504. doi: 10.1080/10705510701301834

Cohrssen, C., and Niklas, F. (2019). Using mathematics games in preschool settings to support the development of children's numeracy skills. *Int. J. Early Years Educ.* 27, 322–339. doi: 10.1080/09669760.2019.1629882

Cohrssen, C., Tayler, C., and Cloney, D. (2014). Playing with maths: implications for early childhood mathematics teaching from an implementation study in Melbourne, Australia. *Education 3-13* 43, 641–652. doi: 10.1080/03004279.2013.848916

Davis, O. S. P., Band, G., Pirinen, M., Haworth, C. M. A., Meaburn, E. L., Kovas, Y., et al. (2014). The correlation between reading and mathematics ability at age twelve has a substantial genetic component. *Nat. Commun.* 5:4204. doi: 10.1038/ncomms5204

De Keyser, L., Bakker, M., Rathé, S., Wijns, N., Torbeyns, J., Verschaffel, L., et al. (2020). No association between the home math environment and numerical and patterning skills in a large and diverse sample of 5- to 6-year-olds. *Front. Psychol.* 11:547626. doi: 10.3389/fpsyg.2020.547626

DeFlorio, L., and Beliakoff, A. (2015). Socioeconomic status and preschoolers' mathematical knowledge: the contribution of home activities and parent beliefs. *Early Educ. Dev.* 26, 319–341. doi: 10.1080/10409289.2015.968239

del Río, M. F., Strasser, K., Cvencek, D., Susperreguy, M. I., and Meltzoff, A. N. (2019). Chilean kindergarten children's beliefs about mathematics: family matters. *Dev. Psychol.* 55, 687–702. doi: 10.1037/dev0000658

del Río, M. F., Susperreguy, M. I., Strasser, K., Cvencek, D., Iturra, C., Gallardo, I., et al. (2020). Early sources of children's math achievement in chile: the role of parental beliefs and feelings about math. *Early Educ. Dev.* 32, 637–652. doi: 10.1080/10409289.2020.1799617

del Río, M. F., Susperreguy, M. I., Strasser, K., and Salinas, V. (2017). Distinct influences of mothers and fathers on kindergartners' numeracy performance: the role of math anxiety, home numeracy practices, and numeracy expectations. *Early Educ. Dev.* 28, 939–955. doi: 10.1080/10409289.2017.1331662

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Dev. Psychol.* 43, 1428–1446. doi: 10.1037/0012-1649.43.6.1428

Eccles, J. S., Adler, T. F., Futtermann, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). "Expectancies, values, and academic behaviors," in *Achievement and Achievement Motivation*, ed. J. T. Spence (New York, NY: W. H. Freeman), 75–146.

Ehlert, A., Ricken, G., and Fritz, A. (2020). *MARKO-Screening - Mathematik- und Rechenkonzepte im Vorschulalter - Screening [MARKO-Screening - Mathematics and Concepts of Calculation before School Entry]*. Göttingen: Hogrefe.

Elliott, L., and Bachman, H. J. (2018). How do parents foster young children's math skills? *Child Dev. Perspect.* 12, 16–21. doi: 10.1111/cdep.12249

Endlich, D., Berger, N., Küspert, P., Lenhard, W., Marx, P., Weber, J., et al. (2017). *WVT: Würzburger Vorschultest: Erfassung schriftsprachlicher und mathematischer (Vorläufer-) Fertigkeiten und sprachlicher Kompetenzen im letzten Kindergartenjahr [WVT: Würzburg preschool test: Assessment of literacy and mathematical (precursor) abilities and linguistic competencies in the last year of kindergarten]*. Göttingen: Hogrefe.

Gasteiger, H., and Moeller, K. (2021). Fostering early numerical competencies by playing conventional board games. *J. Exp. Child Psychol.* 204:105060. doi: 10.1016/j.jecp.2020.105060

Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., and Nugent, L. (2009). First-grade predictors of mathematical learning disability: a latent class trajectory analysis. *Cogn. Dev.* 24, 411–429. doi: 10.1016/j.cogdev.2009.10.001

Gould, P. (2012). What number knowledge do children have when starting Kindergarten in NSW? *Aust. J. Early Child.* 37, 105–110. doi: 10.1177/183693911203700314

Gunderson, E. A., Ramirez, G., Levine, S. C., and Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles* 66, 153–166. doi: 10.1007/s11199-011-9996-2

Hart, S. A., Little, C., and van Bergen, E. (2021). Nurture might be nature: cautionary tales and proposed solutions. *NPJ Sci. Learn.* 6, 1–12. doi: 10.1038/s41539-020-00079-z

Hart, S. A., Petrill, S. A., Thompson, L. A., and Plomin, R. (2009). The ABCs of math: a genetic analysis of mathematics and its links with reading ability and general cognitive ability. *J. Educ. Psychol.* 101, 388–402. doi: 10.1037/a0015115

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., and Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Pers. Soc. Psychol. Bull.* 31, 1369–1385. doi: 10.1177/0146167205275613

Hornburg, C. B., Borriello, G. A., Kung, M., Lin, J., Litkowski, E., Cosso, J., et al. (2021). Next directions in measurement of the home mathematics environment: an international and interdisciplinary perspective. *J. Numer. Cogn.* 7, 195–220. doi: 10.5964/jnc.6143

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equat. Model.* 6, 1–55. doi: 10.1080/10705519909540118

IBM Corp (2021). *IBM SPSS Statistics for Windows (Version 28) [Computer software]*. Armonk, NY: IBM Corp.

Jacobs, J. E., Davis-Kean, P., Bleeker, M., Eccles, J. S., and Malanchuk, O. (2005). "I can, but I don't want to": the impact of parents, interests, and activities on gender differences in math," in *Gender Differences in Mathematics: An Integrative Psychological Approach*, eds A. M. Gallagher and J. C. Kaufman (Cambridge: Cambridge University Press).

JASP Team (2021). *JASP (Version 0.16.0.0) [Computer software]*. Available online at: https://jasp-stats.org/download/ [accessed on February 16, 2022].

Jordan, N. C., Glutting, J., and Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learn. Individ. Differ.* 20, 82–88. doi: 10.1016/j.lindif.2009.07.004

Jordan, N. C., Kaplan, D., Locuniak, M. N., and Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learn. Disabil. Res. Pract.* 22, 36–46. doi: 10.1111/j.1540-5826.2007.00229.x

Jordan, N. C., Kaplan, D., Nabors Oláh, L., and Locuniak, M. N. (2006). Number sense growth in kindergarten: a longitudinal investigation of children at risk for mathematics difficulties. *Child Dev.* 77, 153–175. doi: 10.1111/j.1467-8624.2006.00862.x

Jordan, N. C., Kaplan, D., Ramineni, C., and Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. *Dev. Psychol.* 45, 850–867. doi: 10.1037/a0014939

Kaya, S., and Lundeen, C. (2010). Capturing parents' individual and institutional interest toward involvement in science education. *J. Sci. Teach. Educ.* 21, 825–841.

Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., and Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. *NPJ Sci. Learn.* 3:12. doi: 10.1038/s41539-018-0028-7

Kleemans, T., Peeters, M., Segers, E., and Verhoeven, L. (2012). Child and home predictors of early numeracy skills in kindergarten. *Early Child. Res. Q.* 27, 471–477. doi: 10.1016/j.ecresq.2011.12.004

Krajewski, K. (2018). *MBK 0. Test Mathematischer Basiskompetenzen im Kindergartenalter [MBK 0. Assessment of basic Mathematical Competencies in Kindergarten age]*, 1st Edn. Göttingen: Hogrefe.

Krajewski, K., and Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: findings from a four-year longitudinal study. *Learn. Instruct.* 19, 513–526. doi: 10.1016/j.learninstruc.2008.10.002

LeFevre, J. A., Polyzoi, E., Skwarchuk, S. L., Fast, L., and Sowinski, C. (2010). Do home numeracy and literacy practices of Greek and Canadian parents predict the numeracy skills of kindergarten children? *Int. J. Early Years Educ.* 18, 55–70. doi: 10.1080/09669761003693926

LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B., Fast, L., Kamawar, D., and Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Can. J. Behav. Sci.* 41, 55–66.

Lewis, M., and Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nat. Hum. Behav.* 4, 1021–1028. doi: 10.1038/s41562-020-0918-6

Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychol. Methods* 21, 369–387. doi: 10.1037/met0000093

Li, C.-H. (2021). Statistical estimation of structural equation models with a mixture of continuous and categorical observed variables. *Behav. Res. Methods* 53, 2191–2213. doi: 10.3758/s13428-021-01547-z

Lionetti, F., Keijsers, L., Dellagiulia, A., and Pastore, M. (2016). Evidence of factorial validity of parental knowledge, control and solicitation, and adolescent disclosure scales: when the ordered nature of likert scales matters. *Front. Psychol.* 7:941. doi: 10.3389/fpsyg.2016.00941

Ludwig, K. U., Sämann, P., Alexander, M., Becker, J., Bruder, J., Moll, K., et al. (2013). A common variant in myosin-18B contributes to mathematical abilities in children with dyslexia and intraparietal sulcus variability in adults. *Transl. Psychiatry* 3:e229. doi: 10.1038/tp.2012.148

Melhuish, E., Ereky-Stevens, K., Ariescu, A., Penderi, E., Rentzou, K., Tawell, A., et al. (2015). *A Review of Research on the Effects of Early Childhood Education and Care (ECEC) upon child development: CARE project; Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC)*. Available online at: file:///C:/Users/ra29coz/AppData/Local/Temp/new_version_CARE_WP4_D4_1_Review_on_the_effects_of_ECEC.pdf [accessed on December 11, 2016].

Meunier, J.-C., and Roskam, I. (2009). Self-efficacy beliefs amongst parents of young children: validation of a self-report measure. *J. Child Fam. Stud.* 18, 495–511. doi: 10.1007/s10826-008-9252-8

Miller, D. I., Eagly, A. H., and Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: evidence from 66 nations. *J. Educ. Psychol.* 107, 631–644. doi: 10.1037/edu0000005

Missall, K., Hojnoski, R. L., Caskie, G. I. L., and Repasky, P. (2015). Home numeracy environments of preschoolers: examining relations among mathematical activities, parent mathematical beliefs, and early mathematical skills. *Early Educ. Dev.* 26, 356–376. doi: 10.1080/10409289.2015.968243

Missall, K. N., Hojnoski, R. L., and Moreano, G. (2016). Parent–child mathematical interactions: examining self-report and direct observation. *Early Child Dev. Care* 187, 1896–1908. doi: 10.1080/03004430.2016.1193731

Morey, R. D., and Rouder, I. N. (2015). *Computation of Bayes Factors for Common Designs [R package BayesFactor version 0.9.11-1.]. Comprehensive R Archive Network (CRAN)*. Available Online at: https://cran.r-project.org/web/packages/BayesFactor/index.html [accessed on December 13, 2021].

Mues, A., Birtwistle, E., Wirth, A., and Niklas, F. (2021). Parental (STEM) occupations, the home numeracy environment, and kindergarten children's numerical competencies. *Educ. Sci.* 11:819. doi: 10.3390/educsci11120819

Muthen, L. K., and Muthen, B. O. (2021). *MPlus User's Guide*, 8th Edn. Los Angeles, CA: Muthén & Muthén.

Napoli, A. R., and Purpura, D. J. (2018). The home literacy and numeracy environment in preschool: cross-domain relations of parent-child practices and child outcomes. *J. Exp. Child Psychol.* 166, 581–603. doi: 10.1016/j.jecp.2017.10.002

Navarro, M. G., Braham, E. J., and Libertus, M. E. (2018). Intergenerational associations of the approximate number system in toddlers and their parents. *Br. J. Dev. Psychol.* 36, 521–539. doi: 10.1111/bjdp.12234

Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., et al. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Child. Res. Q.* 36, 550–560. doi: 10.1016/j.ecresq.2016.02.003

Niepel, C., Stadler, M., and Greiff, S. (2019). Seeing is believing: gender diversity in STEM is related to mathematics self-concept. *J. Educ. Psychol.* 111, 1119–1130. doi: 10.1037/edu0000340

Niklas, F., Annac, E., and Wirth, A. (2020a). App-based learning for kindergarten children at home (Learning4Kids): study protocol for cohort 1 and the kindergarten assessments. *BMC Pediatr.* 20:554. doi: 10.1186/s12887-020-02432-y

Niklas, F., Cohrssen, C., and Tayler, C. (2016). Improving preschoolers' numerical abilities by enhancing the home numeracy environment. *Early Educ. Dev.* 27, 372–383. doi: 10.1080/10409289.2015.1076676

Niklas, F., and Schneider, W. (2012). Die Anfänge geschlechtsspezifischer Leistungsunterschiede in mathematischen und schriftsprachlichen Kompetenzen [The beginning of gender-based performance differences in mathematics and linguistic competencies]. *Z. Entwicklungspsychol. Pädagog. Psychol.* 44, 123–138. doi: 10.1026/0049-8637/a000064

Niklas, F., and Schneider, W. (2014). Casting the die before the die is cast: the importance of the home numeracy environment for preschool children. *Eur. J. Psychol. Educ.* 29, 327–345. doi: 10.1007/s10212-013-0201-6

Niklas, F., and Schneider, W. (2017). Home learning environment and development of child competencies from kindergarten until the end of elementary school. *Contemp. Educ. Psychol.* 49, 263–274. doi: 10.1016/j.cedpsych.2017.03.006

Niklas, F., Wirth, A., Guffler, S., Drescher, N., and Ehmig, S. C. (2020b). The home literacy environment as a mediator between parental attitudes toward shared reading and children's linguistic competencies. *Front. Psychol.* 11:1628. doi: 10.3389/fpsyg.2020.01628

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., et al. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proc. Natl. Acad. Sci. U. S. A.* 106, 10593–10597. doi: 10.1073/pnas.0809921106

Overstall, A. M., and King, R. (2014). Conting: an r package for bayesian analysis of complete and incomplete contingency tables. *J. Stat. Softw.* 58, 1–27. doi: 10.18637/jss.v058.i07

Peacock-Chambers, E., Martin, J. T., Necastro, K. A., Cabral, H. J., and Bair-Merritt, M. (2017). The influence of parental self-efficacy and perceived control on the home learning environment of young children. *Acad. Pediatr.* 17, 176–183. doi: 10.1016/j.acap.2016.10.010

Pettigrew, K. A., Fajutrao Valles, S. F., Moll, K., Northstone, K., Ring, S., Pennell, C., et al. (2015). Lack of replication for the myosin-18B association with mathematical ability in independent cohorts. *Genes Brain Behav.* 14, 369–376. doi: 10.1111/gbb.12213

Puglisi, M. L., Hulme, C., Hamilton, L. G., and Snowling, M. J. (2017). The home literacy environment is a correlate, but perhaps not a cause, of variations in children's language and literacy development. *Sci. Stud. Read.* 21, 498–514. doi: 10.1080/10888438.2017.1346660

Saracho, O. N., and Spodek, B. (2008). Fathers: the 'invisible' parents. *Early Child Dev. Care* 178, 821–836. doi: 10.1080/03004430802352244

Schmader, T., Johns, M., and Barquissau, M. (2004). The costs of accepting gender differences: the role of stereotype endorsement in women's experience in the math domain. *Sex Roles* 50, 835–850. doi: 10.1023/B:SERS.0000029101.74557.a0

Schumacker, R. E., and Lomax, R. G. (2010). *A Beginner's Guide to Structural Equation Modeling*, 3. Edn. Oxfordshire: Routledge.

Skwarchuk, S.-L. (2009). How do parents support preschoolers' numeracy learning experiences at home? *Early Child. Educ. J.* 37, 189–197. doi: 10.1007/s10643-009-0340-1

Skwarchuk, S.-L., Sowinski, C., and LeFevre, J.-A. (2014). Formal and informal home learning activities in relation to children's early numeracy and literacy skills: the development of a home numeracy model. *J. Exp. Child Psychol.* 121, 63–84. doi: 10.1016/j.jecp.2013.11.006

Sonnenschein, S., Galindo, C., Metzger, S. R., Thompson, J. A., Huang, H. C., and Lewis, H. (2012). Parents' beliefs about children's math development and children's participation in math activities. *Child Dev. Res.* 2012, 851657. doi: 10.1155/2012/851657

Sonnenschein, S., Metzger, S. R., and Thompson, J. A. (2016). Low-income parents' socialization of their preschoolers' early reading and math skills. *Res. Hum. Dev.* 13, 207–224. doi: 10.1080/15427609.2016.1194707

Susperreguy, M. I., Douglas, H., Xu, C., Molina-Rojas, N., and LeFevre, J.-A. (2020). Expanding the home numeracy model to chilean children: relations among parental expectations, attitudes, activities, and children's mathematical outcomes. *Early Child. Res. Q.* 50, 16–28. doi: 10.1016/j.ecresq.2018.06.010

Susperreguy, M. I., Jiménez Lira, C., Xu, C., LeFevre, J.-A., Blanco Vega, H., Benavides Pando, E. V., et al. (2021). Home learning environments of children

in mexico in relation to socioeconomic status. *Front. Psychol.* 12:626159. doi: 10.3389/fpsyg.2021.626159

Tatto, M. T., Schwille, J., Senk, S. L., Ingvarson, L., Rowley, G., Peck, R., et al. (2012). *Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics in 17 countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. Paris: IEA.

Tazouti, Y., and Jarlégan, A. (2019). The mediating effects of parental self-efficacy and parental involvement on the link between family socioeconomic status and children's academic achievement. *J. Fam. Stud.* 25, 250–266. doi: 10.1080/13229400.2016.1241185

Tomasetto, C., Mirisola, A., Galdi, S., and Cadinu, M. (2015). Parents' math–gender stereotypes, children's self-perception of ability, and children's appraisal of parents' evaluations in 6-year-olds. *Contemp. Educ. Psychol.* 42, 186–198. doi: 10.1016/j.cedpsych.2015.06.007

Vasilyeva, M., Laski, E., Veraksa, A., Weber, L., and Bukhalenkova, D. (2018). Distinct pathways from parental beliefs and practices to children's numeric skills. *J. Cogn. Dev.* 19, 345–366. doi: 10.1080/15248372.2018.1483371

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Zippert, E. L., and Rittle-Johnson, B. (2020). The home math environment: more than numeracy. *Early Child. Res. Q.* 50, 4–15. doi: 10.1016/j.ecresq.2018.07.009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Education

ORIGINAL RESEARCH
published: 16 May 2022
doi: 10.3389/feduc.2022.840618

Social Contexts and Gender Disparities in Students' Competence Beliefs: The Role of Gender-Stereotypical Beliefs and Achievement Patterns in the Classroom for Students' Self-Concept in Gender-Stereotypical Subjects

Ida G. Andersen and Emil Smith*

Department of Educational Sociology, Danish School of Education, Aarhus University, Aarhus, Denmark

OPEN ACCESS

Edited by:
Daniela Barni,
University of Bergamo, Italy

Reviewed by:
Marlene Kollmayer,
University of Vienna, Austria
Malte Jansen,
Institute for Educational Quality Improvement (IQB), Germany

****Correspondence:***
Ida G. Andersen
idagran@edu.au.dk

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *21 December 2021*
Accepted: *28 March 2022*
Published: *16 May 2022*

Citation:
Andersen IG and Smith E (2022) Social Contexts and Gender Disparities in Students' Competence Beliefs: The Role of Gender-Stereotypical Beliefs and Achievement Patterns in the Classroom for Students' Self-Concept in Gender-Stereotypical Subjects. Front. Educ. 7:840618. doi: 10.3389/feduc.2022.840618

This study investigated the role of social contexts for gender disparities in education by examining the associations between gender-stereotypical beliefs (GSB) of students, peers, and teachers and gender achievement patterns in the classroom and students' self-concept in language and math. We applied multilevel models with school fixed effects to a unique sample of combined survey and register data from Denmark to analyze detailed learning environments within schools and their correlations with gender differences in self-concept across subject domains. Results showed a gender gap in favor of boys in mathematics, net of academic achievement that were consistent across classrooms. In language, the influence of gender varied across classrooms. Furthermore, although GSB and gender achievement patterns did not alter the gender gap in either language or mathematics, we found that they moderated the relationship between gender and self-concept in heterogeneous ways across subjects. While teachers' GSB increased the gender gap in language by decreasing boys' self-concept, the students' own GSB was more important for students' self-concept in mathematics. Moreover, girls' mathematics self-concept was lower in classrooms, in which, female peers had a relatively higher level of mathematics achievement compared to boys, suggesting that counter-stereotypical achievement patterns in the classroom do not increase students' self-concept in subjects with strong gender stereotypes. On the contrary, girls are most likely to compare themselves to female peers, resulting in a

negative association with self-evaluations. Our results highlighted the role played by social contexts in schools in the generation of gender differences in self-concept in traditionally stereotyped subject domains, but also showed important differences in how boys and girls were affected by their learning environments across different subject domains, suggesting there are different mechanisms at play.

Keywords: gender differences, social contexts, gender-stereotype, self-concept, multilevel (hierarchical) regression

INTRODUCTION

Recent decades have seen considerable change in patterns of gender disparities in education. Two notable tendencies are evident and contribute to a complex pattern of vertical and horizontal gender inequality. On the one hand, women have increased their levels of participation in tertiary education and have now surpassed men in terms of educational attainment, both in the United States (DiPrete and Buchmann, 2013) and in most European countries (Vincent-Lancrin, 2008). Furthermore, studies have shown that female students outperform male students in most subject domains, particularly in language (Voyer and Voyer, 2014; Reilly et al., 2019). However, while female students a few decades ago had almost universally lower academic achievement in STEM (Science, Technology, Engineering, and Math)-related subjects compared to male students (Reilly, 2012), there has been a slight decline in the number of countries with gender achievement gaps in mathematics (Hyde et al., 2008; Else-Quest et al., 2010; OECD, 2019) and the gender gap in science now favors girls in most countries (Neuschmidt et al., 2008; Mullis et al., 2020). On the other hand, despite women's higher levels of vertical educational attainment, strong horizontal gender segregation persists. Accordingly, women remain underrepresented in most STEM fields and particularly within math-intensive fields (UNESCO, 2017; McNally, 2020). Whether horizontal segregation constitutes a problem is open to debate. Nevertheless, research has suggested that horizontal segregation, beginning as early as in upper secondary education, is linked to subsequent inequality in the labor market, with math-intensive tracks providing the greatest advantages (Birkelund et al., 2021).

An important conclusion from prior research is that disparity in academic and educational outcomes do not reflect inherent gender differences, but are a result of culturally embedded gender beliefs in the form of stereotypes—that is, a belief or set of beliefs regarding the characteristics, attributes, or behaviors of a particular group or category of people (Hilton and Von Hippel, 1996). Empirical research has shown that gender disparities in prior academic achievement cannot alone explain gender differences in entry to STEM fields (Riegle-Crumb et al., 2012). Combined with variations in gender disparities in educational outcomes across national contexts and cohorts (Penner, 2008), these findings support sociological theories on the social nature of gender. Although gender can be perceived as a social structure with importance for educational stratification (Risman, 2004; England, 2010), gender is not a fixed category, but a social construction constituted through a multilayered system of macro-level structures and cultural beliefs, as well as micro-level contexts of personal interactions and exchanges (Correll, 2004; Ridgeway and Correll, 2004). Accordingly, it has been suggested that gender is likely to be more salient in some social interactions than in others and different social settings may activate certain stereotypes or social scripts regarding gender identity and achievement (Ridgeway, 2009). So far, research has tended to focus on a specific dimension of gender differences (e.g., boys' lower reading scores or girls' under-representation in STEM), thus, providing important yet narrow explanations for the existence of such differences (Legewie and DiPrete, 2014; Mann et al., 2015; Retelsdorf et al., 2015; Riegle-Crumb and Morton, 2017; Muntoni and Retelsdorf, 2018). Thus, there is a shortage of theoretical explanations covering the broad constellation of gender differences and similarities in terms of educational outcomes, as well as empirical research investigating the social construction of gender identity and inequality in social contexts (Riegle-Crumb et al., 2018). Moreover, we currently have limited knowledge of the presence and variability of cultural gender beliefs and stereotypes across different school contexts characterized by diverse achievement-related gender structures and how they may shape gender disparities in competence beliefs of male and female students.

To address this gap, this paper investigates the role of school contexts in students' competence beliefs in traditionally gendered subjects. Previous research has shown that students' competence beliefs – or *academic self-concept* – are key factors in predicting educational behavior (Wigfield and Eccles, 2000) and documented the existence of gender gaps in self-concept in different academic subjects in line with traditional gendered patterns. While boys typically hold more positive self-concepts in mathematics (Goldman and Penner, 2016) and science (Sikora and Pokropek, 2012), girls typically hold more positive self-concepts in language (Jacobs et al., 2002), net of actual achievement. Building on psychological and sociological research, we focused on two determinants of competence beliefs that potentially generate gendered patterns in attainment. First, we examined the role of gender-stereotypical beliefs (GSB). Girls generally hold lower self-concepts than boys at equal ability levels (Correll, 2001) and it has been suggested that self-concept is an expression of internalized gender beliefs (Charles and Bradley, 2009; Eccles, 2011; Breda et al., 2020). Accordingly, if a girl believes that boys are more competent in mathematics, she might view mathematical competence as inconsistent with female gender identity and thus, doubt her mathematical ability. Indeed, research has shown associations between gender

stereotypes and students' self-concept in the traditionally male-dominated subject of mathematics (Riegle-Crumb and Peng, 2021) and the traditionally female-dominated subject of reading (Retelsdorf et al., 2015; Muntoni et al., 2021). Second, we investigated whether gender-stereotypical achievement patterns in the classroom affected the competence beliefs of male (female) students in mathematics (language) and the extent to which counter-stereotypical achievement reduced gender gaps in self-concept across gender-stereotypical subject domains.

We applied multilevel models to combined survey and register data on 1,099 Danish compulsory school students and their respective language and mathematics teachers to investigate how GSB in the classroom influence students' self-concept in language and mathematics. Our study contributes and adds to the existing body of research in three specific ways. First, unlike most previous research, which has typically focused on gender stereotypes among either students, peers, or teachers, we used a unique data set that includes all three dimensions. Combined with the fact that our data was sampled at the classroom level, we were able to investigate very rich variations in GSB within and across different social contexts in schools. Second, we expand on previous research on GSB in schools by also including information on gender-achievement patterns in classrooms. While gender beliefs are one important dimension of the classroom environment, another is the relative achievement of male/female students across different subject domains. To the best of our knowledge, research on the influence of gender-achievement patterns in the classroom on gender differences has been limited or non-existent. We believe that this is an important dimension of the social context in schools with possible implications for the construction of gender and inequality across classrooms. Third, we added to prior research on the role of peers in terms of students' educational outcomes in general and gender differences in domain-specific self-concept, in particular, by distinguishing between male and female peers. This distinction was motivated by empirical research that has shown that STEM-related outcomes and course-taking in high school are influenced by same-gender friendships (Riegle-Crumb et al., 2006; Raabe et al., 2019).

THEORETICAL PERSPECTIVES

Gender Differences in Students' Competence Beliefs

The expectancy-value theory offers a powerful framework for understanding gender differences in students' achievement-related behavior. According to this perspective, there are two components of students' motivation: expectancy beliefs (e.g., competence beliefs) and task value beliefs (e.g., interest and utility) (Eccles and Wigfield, 1995). Individuals holding higher expectancy and task beliefs are more likely to pursue a specific subject. There is a degree of overlap between expectancy beliefs and the terms self-efficacy and academic self-concept, and expectancy beliefs have often been operationalized as such in empirical studies (Bong and Skaalvik, 2003; Marsh et al., 2019). Academic self-concept refers to an individual's assessment of their ability (Wigfield and Eccles, 2000) and varies across different subject domains, such as language and mathematics (Marsh et al., 2006; Jansen et al., 2014). Students' beliefs about their competencies most likely reflect two distinct elements: a descriptive and a normative element. On the one hand, students' academic self-concept can be expected to reflect their actual ability. On the other hand, students evaluate their ability through a subjective lens, reflecting on social and cultural norms. Previous empirical research has supported this by showing that even when performing at the same level as their male peers, girls were less confident in their math and science abilities (Else-Quest et al., 2010).

Many scholars have investigated gender differences in students' competence beliefs across different subject domains, as well as the sources of these differences. Findings in this field have generally reflected the culturally gendered perception of subject domains, with math as a male domain and language as a female domain, in that girls have a lower academic self-concept than boys in math domains (Goldman and Penner, 2016), while boys have a lower academic self-concept than girls in language domains (Durik et al., 2006; Ireson and Hallam, 2009). Importantly, these gender differences have often remained even when controlling for achievement. In a study using data from TIMSS 2015, Mejía-Rodríguez et al. (2021) documented the existence of a gender gap in mathematics self-concept in most participating countries, usually favoring boys. This difference in self-concept across male and female students remained after controlling for academic achievement and other covariates and was evident as early as 4th grade. However, findings have been less robust about the language self-concept of male students, with some studies finding no statistically significant gender differences (Skaalvik and Skaalvik, 2004; Evans et al., 2011). Nevertheless, boys' self-concept, rather than their innate ability, is an important predictor of their language achievement (Heyder et al., 2017). Furthermore, girls' self-concept in math is related to their belief in math ability as innate (i.e., a fixed mindset), whereas no such relationship has been found between boys' self-concept in language and their beliefs about innate language ability (Heyder et al., 2021), suggesting that girls may, to a larger degree, have internalized beliefs about gender and ability.

School Contexts and the Social Construction of Gender and Competence Beliefs

According to many sociological and psychological theories, students' achievement-related outcomes are influenced by social contexts and interactions. Specifically, sociological scholarship has repeatedly documented the centrality of social influences for gender differences in achievement-related outcomes in general (Legewie and DiPrete, 2012; Salikutluk and Heyne, 2017; Raabe et al., 2019) and gender disparities in orientations toward STEM in particular (Crosnoe et al., 2008; Legewie and DiPrete, 2014; Riegle-Crumb and Morton, 2017). Such research suggests that gender is a multi-level system that is (re-)constructed through interactions and experiences in social contexts, such as schools and classrooms, and that to better understand gender inequality we need to consider variation in

such contexts (Ridgeway and Correll, 2004; Risman, 2004). The production of gender roles and positions in schools is likely a consequence of the fact that images of femininity and masculinity are socially constructed in everyday student interactions with peers and teachers, who, thus, collectively functions as active agents of gender socialization. A similar perspective can be found in Eccles (1994) expectancy-value theory, which states that socializing agents, such as peers and teachers, play an important role in shaping students' academic self-concept. In particular, peers have been found to play an important part in gender role socialization (Witt, 2000) and, perhaps, even discourage gender non-conformity through victimization (Lamb et al., 1980; Aspenlieder et al., 2009). Moreover, same-gender peers are particularly important in shaping gendered patterns in academic outcomes, by both serving to promote gender conformity in educational decisions (Riegle-Crumb et al., 2006; Rosenqvist, 2018; Raabe et al., 2019) and serving as reference points for self-evaluations that shape competence beliefs (Thijs et al., 2010).

Gender-Stereotypical Beliefs in the Classroom

Stereotypes can be defined as reflecting "general expectations about members of particular social groups" (Ellemers, 2018, p. 276). In education, different subject domains have often been shown to be gendered. In general, math and math-related domains are stereotyped as male, while language domains are stereotyped as female (Charles and Bradley, 2009; Martinot et al., 2012; Heyder and Kessels, 2013; Nowicki and Lopata, 2017). Stereotypical expectations can reflect both actual and false differences. On the one hand, descriptive expectations reflect an observed gendered pattern, such as the mathematics achievement of boys and girls in a particular classroom, and, thus, a seemingly "true" representation of gender differences. On the other hand, expectations may reflect not only students' actual achievement but also cultural perceptions of gender and how well-suited boys and girls are for different subject domains. Importantly, even the observed gender differences rarely reflect inherent biological gender differences, but can most often be ascribed to socialization (Ellemers, 2018). Furthermore, expectations concerning an entire social group, such as girls or boys, are often imprecise when directed at a single individual, but can still have negative consequences by reinforcing the stereotype.

The students' own stereotypical beliefs have been shown to have a negative influence on outcomes, such as achievement and self-concept (Plante et al., 2009, 2013; Heyder and Kessels, 2013), but stereotypes often stem from significant others in social contexts. The negative consequences of being exposed to stereotypical beliefs have often been framed as a stereotype threats, which describes how exposure to stereotypical beliefs can lead to stereotypes becoming self-fulfilling prophecies (Steele, 1997). Stereotype threat has been found in controlled experimental settings for both girls in math achievement (Flore and Wicherts, 2015) and boys in language achievement (Pansu et al., 2016). However, recent studies have explored the role of stereotype threat and the generally negative impact of gender-stereotypical beliefs in the context of the classroom, emphasizing the role of the gender-stereotypical beliefs of socializing agents such as teachers and peers for various student outcomes. Specifically, factorial survey studies have found evidence of teachers' judgment of students being biased by gender stereotypes (Holder and Kessels, 2017), while the gender-stereotypical beliefs of teachers have been found to negatively affect girls' achievement (Alan et al., 2018) and self-concept (Heyder et al., 2019) in mathematics and boys' self-concept (Retelsdorf et al., 2015) and achievement (Muntoni and Retelsdorf, 2018) in reading. Similarly, the gender-stereotypical beliefs of peers have been found to negatively affect girls' achievement in math (Salikutluk and Heyne, 2017) and boys' self-concept in reading (Muntoni et al., 2021).

Gender-Stereotypical Achievement Patterns in the Classroom

In addition to culturally embedded perceptions of gender, an important aspect of social contexts in schools is the achievement pattern in the classroom; i.e., the academic performance of a student's peers. To understand how students form academic self-concepts and the role played by social comparison in an educational setting, (Marsh, 1987) proposed the idea of the big-fish-little-pond (BFLP) effect. According to this perspective, students compare their academic ability to that of their classroom peers when forming an academic self-concept. This implies that students in higher-achieving social contexts have lower academic self-concepts than students of similar ability in lower-achieving settings. Numerous empirical studies have supported the BFLP effect (Seaton et al., 2010; Fang et al., 2018; Loyalka et al., 2018), as well as confirmed its generalizability across cultural contexts (Seaton et al., 2010; Loyalka et al., 2018). However, research has also shown that social comparison processes differ across gender and that students' academic self-concept is mainly affected by same-gender classmates and only to a lesser extent by different-gender peers (Thijs et al., 2010). Furthermore, in the mathematics domain, gender has been shown to moderate the BFLP effect since female students tend to be more responsive than male students to the achievement of peers (Plieninger and Dickhäuser, 2015). While previous research has provided important evidence that gender is a relevant frame of reference within classrooms, as well as of heterogeneous influences of social comparison across gender, no study, to date, has investigated how classroom achievement patterns intersect with cultural perceptions of gender and how this influences students' evaluation of their competences. In this paper, we argue that when gender is used as a frame of reference, gender identity plays an important role in the social comparison process. In contrast to the more general BFLP perspective in which students' self-concept is negatively affected by average peer performance, students may be positively affected by high-achieving same-sex peers because such peers can function as role models and, thus, provide counter-stereotypical evidence of gender suitability within a subject domain. Accordingly, girls may hold more positive self-concepts in mathematics if surrounded by female peers that counteract the stereotype that girls have lower ability than boys within this subject domain do. Consequently, counter-stereotypical

achievement patterns potentially can disrupt gender stereotypes and how they influence students' self-concept. We defined counter-stereotypical achievement patterns as math (language) classrooms in which female (male) students have higher average achievement than male (female) students—i.e., classrooms in which girls' (boys') relative achievement contradicts the stereotypical expectation. Specifically, we expected that being in a classroom, in which female peers outperform male peers in mathematics achievement could increase girls' mathematics self-concept through identification.

The Present Study

In this study, we investigated the influence of gender on students' self-concept within and between classroom contexts in language and mathematics, which represents traditionally stereotyped subject domains. We posed three specific research questions:

1 Is there a gender gap in students' self-concept in language and mathematics net of actual subject-specific achievement, and does the influence of gender vary across classrooms?
2 Do students, peers, and teachers' GSB affect gender gaps in self-concept in language and mathematics?
3 Does the gender-achievement pattern in the classroom affects gender gaps in language and mathematics self-concept?

First, based on previous research on gender gaps in education, we hypothesized that there is a gender gap in self-concept in the language (Durik et al., 2006; Ireson and Hallam, 2009) and mathematics (Goldman and Penner, 2016). Specifically, we expected girls to hold a more positive self-concept in language and boys in mathematics, net of actual achievement in the respective subjects.

> **H1:** There is a gender gap in students' self-concept in language and mathematics, following traditional gender-stereotypical patterns:
>
> **H1a:** Girls hold more positive self-concepts compared to boys in language net of actual achievement.
>
> **H1b:** Boys hold more positive self-concepts compared to girls in mathematics net of actual achievement.

Second, drawing on previous research on gender stereotypes across subject domains, we hypothesized that GSB affects gender gaps in the language (Retelsdorf et al., 2015; Muntoni et al., 2021) and mathematics (Heyder et al., 2019). Contrary to most previous research, our data included information on the GSB of students themselves as well as their peers and (subject-specific) teachers. Furthermore, due to the sampling of full classrooms, we were able to distinguish between male and female peers to examine the extent to which the influence of the GSB on same- and different-gender peers differ. Specifically, we hypothesized that GSB about the language of students, peers, and teachers would be associated with a more positive language self-concept for girls and less positive for boys. By contrast, we expected that GSB about the mathematics of students, peers, and teachers would be associated with a more positive mathematics self-concept for boys and less positive for girls.

> **H2:** GSB in the classroom context influences gender gaps in language and mathematics:
>
> **H2a:** GSB, regarding language, is associated with more positive language self-concepts for girls and less positive for boys, net of actual achievement.
>
> **H2b:** GSB, regarding mathematics, is associated with more positive mathematics self-concepts for boys and less positive for girls, net of actual achievement.

Third, while one dimension of the classroom gender context is GSB, another is the specific gender-achievement pattern. Does subject-specific achievement in the classroom follow traditional gender-stereotypical patterns or is there a counter-stereotypical achievement pattern with boys (girls) outperforming girls (boys) in the language (mathematics) and with what implications for students' self-concept? Specifically, we hypothesized that counter-stereotypical achievement patterns in the classrooms could alter gender gaps in self-concept.

> **H3:** The specific gender-achievement pattern in the classroom influences gender gaps in language and mathematics:
>
> **H3a:** Counter-stereotypical achievement in mathematics classrooms (i.e., girls outperforming boys) reduces the gender gap by increasing mathematics self-concepts among female students.
>
> **H3b:** Counter-stereotypical achievement in language classrooms (i.e., boys outperforming girls) reduces the gender gap by increasing language self-concepts among male students.

MATERIALS AND METHODS

Data and Sample

The survey was conducted in 2019 as part of the project *Exploring School Culture* (ESCU) at Aarhus University and included a sample of students nested in 94 classrooms at 33 Danish schools, specifically 6th (N = 1,094, response rate = 80.2%) and 9th-grade students (N = 892, response rate = 74.3%) and their teachers in Danish and mathematics (N = 143, response rate = 61.6%). Recruited schools registered their 6th and 9th-grade classrooms, and data were collected among the entire classrooms of students (for more information on recruitment and data collection, see Authors 2021). Through unique student identifiers, we were able to combine the survey data with high-quality background information on parental education, income, and occupation, as well as standardized test scores in reading and mathematics from tests conducted in the 6th and 8th grades for the 6th and 9th-grade students, respectively, from the Danish registers.

Combining survey and register data had four major advantages. First, sampling entire classrooms allowed us to analyze fine-grained gender-specific learning environments that

include all students within a classroom linked to their teachers in language and mathematics, respectively, which represent two important and traditionally gender-stereotyped subject domains. Second, the data included measures of GSB at the student, classroom, and teacher levels across language and mathematics, developed from existing scales (e.g., Martinot et al., 2012) and adjusted to the Danish context, as well as detailed information on students' achievement in the same subjects. Accordingly, contrary to most previous research on gender beliefs and stereotypes in schools, which typically focus on either students, peers, or teachers, we were able to analyze a multilevel dataset that includes all three dimensions. Third, drawing on register data allowed us to include an extensive set of control variables that were not based on students' self-reports and were available for the entire population. Fourth, linking survey data on students' academic self-concept to information on their actual achievement from national tests enabled us to investigate students' self-concept in different subject domains net of actual achievement in these subjects.

Students' Self-Concept

We measured self-concept in the language and mathematics domains based on two separate four-item scales (e.g., "I am just not good at mathematics/Danish") answered on a five-point Likert scale. These items stemmed from the self-concept measure from the Programme for International Student Assessment 2012 (OECD, 2013), reworked to also fit the Danish language domain. The scales exhibited high internal consistency, as indicated by Cronbach's alpha for both mathematics (9th-grade $\alpha = 0.93$ and 6th-grade $\alpha = 0.90$) and language (9th-grade $\alpha = 0.90$ and 6th-grade $\alpha = 0.87$). See **Appendix Table 1** for the full list of items. In our analyses, we estimated the gender gap in self-concept as the coefficient for the gender variable in a model predicting the self-concept.

Independent Variables

Student-Level Variables

We used test scores from nationally administered mandatory high-stakes standardized tests to measure student achievement in both subjects. The tests were adaptive; i.e., items were adapted to match students' competence level based on previous responses. For 6th-grade students, the test was taken in the spring of 2019, at the same time as the survey was conducted, while for the 9th-grade students, the test was taken the previous year when they were in 8th grade.

The measures of domain-specific GSB were based on items about perceived differences in ability and attitudes from the "mathematics as a gendered domain" scale (Leder and Forgasz, 2002). Items were translated to Danish and reworked to present general statements that respondents could connect to either boys or girls. Specifically, respondents reacted to seven statements (e.g., "The weakest students in mathematics/Danish are. . .") by indicating whether this is most true for girls or boys on an 11-point scale with "primarily girls" at the one extreme (0) and "primarily boys" at the other (10), with a neutral category in the middle (5). The broad scope of the measure sought to capture beliefs that go beyond observed gender differences in achievement. The stereotype scales displayed high internal consistency, with Cronbach's alpha for mathematics (9th grade $\alpha = 0.89$ and 6th grade $\alpha = 0.85$) and language (9th grade $\alpha = 0.92$ and 6th grade $\alpha = 0.87$). For the analyses, both GSB measures were coded, so that a higher score indicated beliefs in line with the common stereotype (i.e., favoring boys in mathematics and girls in language). See **Appendix Table 2** for the full list of items.

Classroom-Level Variables

We included several explanatory variables at the classroom level based on teacher and peer measures. For teachers, we included a measure of GSB corresponding to the one described for students above. Specifically, this measure was based on seven corresponding items that were answered in the same way as for the students. The internal consistency was acceptable for mathematics ($\alpha = 0.71$) and high for language ($\alpha = 0.92$).

For the peer group, we included measures of GSB, mean achievement, and differences in achievement between female and male peers. The peer GSB measure was calculated as the mean of the student GSB scales at the classroom level for each subject and male and female students separately. Peer achievement was calculated as the mean achievement of students at the classroom level. The individual student was excluded when calculating peer variables, to avoid including the student in their peer group.

Finally, we calculated the gender-specific achievement pattern in the classroom using the difference between female and male achievement in mathematics (female achievement minus male achievement), meaning that a positive score on this variable indicated that the female peer group outperformed the male peer group in the classroom. For language, we did the opposite (i.e., male achievement minus female achievement), meaning that a positive score indicated that the male peer group outperformed the female peer group. Controlling for the classroom-specific gender difference in achievement also served to remove residual variance in the GSB measures that simply reflected the observed gender-achievement patterns in the classroom.

Control Variables

As previous research has indicated an association between socioeconomic status (SES) and GSB (Davis, 2007; Cotter et al., 2011; Pampel, 2011), we controlled for SES at both the individual and classroom levels. We measured the SES of the student's father and mother individually through a composite measure of income (in quartiles), years of education, and occupation (four categories: self-employed, employed, student, and unemployed). We combined these measures by deriving polychoric factor scores (Holgado-Tello et al., 2010). For the SES of both father and mother, we aggregated at the classroom level to calculate a measure of peer SES, while excluding the individual student, as for the other peer variables.

To control for spurious associations with teacher GSB, we included controls for teacher characteristics: gender, age in years, education (a dummy indicating whether they hold a master's degree), and teaching experience in years. **Table 1** presents an overview of unstandardized variables included in the analyses. For descriptive statistics by gender, see **Appendix Tables 3**, **4**.

TABLE 1 | Descriptive statistics.

	Language (N students = 1097, teachers = 73)				Mathematics (N students = 934, teachers = 57)			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Student variables								
Self-concept	2.459	0.858	−0.200	3.800	2.229	1.045	−0.200	3.800
Achievement[a]	1.012	0.868	−2.854	4.139	0.559	1.137	−3.369	4.766
Female	0.500	0.500	0.000	1.000	0.499	0.500	0.000	1.000
Student GSB[a]	6.516	0.644	3.000	8.000	4.943	1.522	0.000	10.000
Peer variables								
GSB, male peers[b]	6.535	0.216	5.877	7.211	5.087	1.141	1.839	7.939
GSB, female peers[b]	6.526	0.247	5.714	7.152	4.741	0.918	2.556	7.310
Difference in female/male achievement[b]	−0.146	0.429	−1.220	0.967	0.104	0.546	−1.197	1.467
SES, father[b]	2.712	0.755	0.963	4.423	2.757	0.758	0.963	4.423
SES, mother[b]	2.915	0.837	0.996	4.572	2.912	0.828	0.996	4.572
Teacher variables								
GSB, teacher[a]	3.610	1.398	0.375	7.500	3.765	0.503	2.000	4.750
Female	0.740	0.442	0.000	1.000	0.614	0.491	0.000	1.000
Age	45.301	10.729	19.000	63.000	45.509	10.652	26.000	65.000
Experience	16.973	10.550	1.000	38.000	16.430	10.821	1.000	42.000
Master's degree	0.082	0.277	0.000	1.000	0.070	0.258	0.000	1.000

GSB are measured on a scale ranging from 0 to 10, with higher values implying beliefs by the domain-specific stereotype (higher values favor boys in math and girls in language). [a]Variable is standardized in the empirical analysis. [b]Calculated from the standardized individual-level variable.

Notably, teacher GSB was generally lower than student and peer GSB, suggesting that either teacher were less inclined to hold stereotypical beliefs or that teachers were more inclined to social desirability bias than were the students. Nevertheless, the GSB of teachers generally showed a statistically significant positive correlation with those of students and peers (see **Appendix Tables 5**, **6** in the Appendix for correlation matrices). Notably, however, female peers' GSB and teacher GSB in language were negatively correlated, suggesting that the girls held lower GSB in language, regardless of the GSB of their teacher.

Analytical Strategy

To test our hypotheses, we analyzed data using a linear multilevel regression model consisting of two levels: students (level 1) and classrooms (level 2). The multilevel framework had several advantages in this context. First, the multilevel strategy allowed us to obtain unbiased standard errors from our hierarchical data, despite violating the principle of independent sampling of observations. Second, we were able to partition the variance in self-concept into student-level and classroom-level variance components, allowing us to determine and study the variance that can be attributed to the social context of the classroom. Third, we tested whether the effect of gender on self-concept varied across the classroom context; i.e., if the impact of gender on a student's competence beliefs depended on the social context of the classroom. We examined this for both language and mathematics self-concept by fitting a random slope for gender and testing this more elaborate model against a simpler random intercept model using a likelihood ratio test.

Finally, while our data contained three levels (students nested in classrooms that were nested in schools), our research questions focused on the first two levels. Therefore, we treated the school level as incidental in our models by adding dummy indicators for each school. This corresponded to a school fixed effects approach, which has previously been used in similar analyses (McNeish and Wentzel, 2017). Besides being a strategy for modeling the school level, the fixed effects approach had the added benefit of significantly reducing selection bias by taking into account selection processes in schools, which may otherwise have biased our estimates. This strategy is valid under the assumption that there is the limited systematic selection of students into classrooms within schools—i.e., that classrooms are formed more or less at random (for a similar argument see Ammermueller and Pischke, 2009). Although we cannot completely rule out the selection at the school level, the strategy is realistic in the Danish context, where schools sought to create equal classrooms by "balancing" resources. This particular characteristic of the Danish compulsory school system implies very limited selection within schools and, at worst, a selection process that can be characterized as negative; i.e., "better" and more experienced teachers tend to be allocated to more disadvantaged classrooms, which would yield conservative estimates (for elaboration see Andersen and Reimer, 2019).

All models were estimated separately for each subject. For the analyses, we group-mean centered all student-level variables and grand-mean centered all classroom-level variables. All analyses were carried out using Stata version 17.

Missing values in our data occurred mainly at the teacher level due to non-response. For language, teacher non-response reduced our sample by 17% and for mathematics by 36%. Due to teacher non-response, we had no information to impute teacher variables. Therefore, we chose a listwise deletion approach to handle missing values. We also excluded classrooms with less than eight student responses. This resulted in an

analytical sample of 1,097 students for language and 934 students for mathematics.

RESULTS

We present results from our empirical analysis in three steps. First, we investigated whether or not students' self-concept differed across classrooms and if there was a gender gap in students' self-concept in language and mathematics net of actual academic achievement. Second, we tested whether the influence of gender on students' self-concept varied across language and mathematics classrooms. Third, we introduced measures of gender-stereotypical beliefs and achievement patterns and their interaction with students' gender.

Tables 2, **3** present the impact of student, peer, and teacher covariates on students' self-concept in language and mathematics. In Model 1, we estimated a baseline model including only students' gender and achievement as predictors of self-concept to estimate the mean gender difference in self-concept across language and mathematics. In Model 2, we included school fixed effects to account for unobserved factors at the school level and thus, focus the analysis on within-school variation. Models 3–5 introduced covariates at the student, peer, and teacher levels, respectively, measuring gender-stereotypical beliefs and achievement patterns, as well as their interaction with students' gender. The main influence of these factors indicated how they affected students' self-concept on average—i.e., across both male and female students—and whether including them in the models affected the gender gap in self-concept in language/mathematics, as well as the difference in gender gaps across classrooms. We included cross-level interactions between students' gender and measures of gender-stereotypical beliefs and achievement to assess heterogeneous effects across male and female students. Finally, Model 6 included the full model with all cross-level interactions.

TABLE 2 | Results from a multilevel model of self-concept in language.

	(1)	(2)	(3)	(4)	(5)	(6)
Fixed part						
Student covariates						
Achievement	0.391*** (0.032)	0.390*** (0.031)	0.395*** (0.033)	0.412*** (0.033)	0.417*** (0.032)	0.418*** (0.033)
Female	0.101 (0.062)	0.095 (0.064)	0.093 (0.063)	0.095 (0.064)	0.098 (0.061)	0.104 (0.061)
Student GSB			0.003 (0.050)	0.041 (0.030)	0.041 (0.029)	0.005 (0.047)
Peer covariates						
Mean achievement				0.117** (0.040)	0.167*** (0.045)	0.165*** (0.045)
Male peer GSB				-0.025 (0.051)	-0.030 (0.038)	-0.005 (0.053)
Female peer GSB				-0.022 (0.062)	-0.037 (0.044)	-0.071 (0.062)
Relative female/male achievement				0.042 (0.056)	0.028 (0.044)	0.034 (0.058)
Teacher covariates						
Teacher GSB					-0.146** (0.055)	-0.153** (0.057)
Cross-level interactions						
Female*student GSB			0.077 (0.056)			0.064 (0.061)
Female*female peer GSB				0.022 (0.075)		0.067 (0.073)
Female*male peer GSB				-0.032 (0.067)		-0.061 (0.068)
Female*teacher GSB					0.133* (0.059)	0.158* (0.064)
Female*relative Female/male achievement				-0.038 (0.071)		-0.017 (0.069)
Intercept	0.0482 (0.147)	-0.0970 (0.274)	-0.138 (0.156)	-0.320 (0.284)	-0.429 (0.290)	-0.453 (0.291)
Random part						
Var(classroom)	0.079 (0.031)	0.048 (0.028)	0.048 (0.031)	0.040 (0.025)	0.026 (0.022)	0.025 (0.022)
Var(student)	0.764 (0.035)	0.759 (0.034)	0.754 (0.035)	0.748 (0.033)	0.746 (0.033)	0.744 (0.033)
Var(female)	0.058 (0.036)	0.080 (0.048)	0.083 (0.043)	0.082 (0.047)	0.059 (0.043)	0.057 (0.043)
ICC	0.094	0.059	0.061	0.051	0.034	0.034
-2 LL	-1446.1	-1420.5	-1417.0	-1411.6	-1406.0	-1404.6
AIC/BIC	2908.2/2948.2	2908.9/3078.9	2902.1/3072.1	2913.1/3138.1	2908.0/3148.0	2913.2/3173.2
Student controls	No	No	Yes	Yes	Yes	Yes
Peer controls	No	No	No	Yes	Yes	Yes
Teacher controls	No	No	No	No	Yes	Yes
N(students)	1,097	1,097	1,097	1,097	1,097	1,097
N(classrooms)	73	73	73	73	73	73

*Random slope across classrooms. School's fixed effects. Parameter estimates with robust standard errors in parenthesis. Models estimated by maximum likelihood and robust standard errors. School dummy variables are not presented in the table. Student-level variables are group-mean centered and classroom-level variables are grand-mean centered. *p < 0.05, **p < 0.01, ***p < 0.001.*

TABLE 3 | Results from a multilevel model of self-concept in mathematics.

	(1)	(2)	(3)	(4)	(5)	(6)
Fixed part						
Student covariates						
Achievement	0.560*** (0.030)	0.564*** (0.028)	0.564*** (0.031)	0.611*** (0.036)	0.604*** (0.032)	0.605*** (0.035)
Female	-0.391*** (0.048)	-0.390*** (0.051)	-0.380*** (0.047)	-0.385*** (0.050)	-0.374*** (0.047)	-0.391*** (0.050)
Student GSB			0.197*** (0.050)	0.080 (0.041)	0.075 (0.041)	0.184*** (0.053)
Peer covariates						
Mean achievement				0.229*** (0.045)	0.249*** (0.043)	0.236*** (0.043)
Male peer GSB				-0.141 (0.086)	-0.141* (0.062)	-0.125 (0.087)
Female peer GSB				0.239** (0.087)	0.117 (0.070)	0.169 (0.096)
Relative female/male achievement				0.061 (0.056)	-0.085 (0.048)	0.005 (0.068)
Teacher covariates						
Teacher GSB					-0.027 (0.045)	-0.014 (0.041)
Cross-level interactions						
Female*student GSB			-0.283** (0.096)			-0.270** (0.093)
Female*male peer GSB				0.054 (0.131)		0.025 (0.131)
Female*female peer GSB				-0.183 (0.111)		-0.201 (0.111)
Female*teacher GSB					0.043 (0.043)	0.025 (0.032)
Female*relative female/male achievement				-0.165** (0.062)		-0.178** (0.066)
Intercept	0.387 (0.200)	0.301 (0.255)	0.308 (0.224)	0.127 (0.168)	0.114 (0.190)	0.172 (0.206)
Random part						
Var(classroom)	0.058 (0.016)	0.010 (0.008)	0.011 (0.009)	6.54e-22 (4.41e-20)	5.94e-25 (3.24e-23)	6.79e-25 (3.42e-23)
Var(student)	0.584 (0.025)	0.583 (0.028)	0.567 (0.025)	0.565 (0.025)	0.564 (0.024)	0.551 (0.024)
ICC	0.090	0.017	0.019	1.16e-21	1.05e-24	1.23e-24
-2LL	-1101.5	-1080.1	-1068.4	-1058.3	-1057.6	-1046.6
AIC/BIC	2215.1/2244.1	2220.3/2365.5	2186.8/2307.8	2180.7/2335.5	2185.1/2354.5	2169.2/2353.1
Student controls	No	No	Yes	Yes	Yes	Yes
Peer controls	No	No	No	Yes	Yes	Yes
Teacher controls	No	No	No	No	Yes	Yes
N(students)	934	934	934	934	934	934
N(classrooms)	57	57	57	57	57	57

*Random intercept across classrooms. School's fixed effects. Parameter estimates with robust standard errors in parenthesis. Models estimated by maximum likelihood and robust standard errors. School dummy variables are not presented in the table. Student-level variables are group-mean centered and classroom-level variables are grand-mean centered. *p < 0.05, **p < 0.01, ***p < 0.001.*

Tables 2, **3** show that in both language and mathematics, the ICC was 9%. Accordingly, 9% of the variation in students' self-concept could be attributed to factors at the classroom level. Model 2 showed that there is a gender gap in students' self-concept in mathematics, while the gender gap in language did not reach conventional levels of statistical significance. In mathematics, there was a large gender gap of 0.39 standard deviations in favor of boys. Accordingly, on average, girls tended to evaluate their mathematics competencies as worse than boys, net of their actual academic achievement. In language, the gender gap was reversed: on average, girls evaluated themselves more positively than boys ($b = 0.101$), although this difference was not statistically significant.

To examine whether the influence of gender varied across classrooms, we tested a random slope of gender (i.e., whether the gender gap varied across classrooms) in both subjects using a likelihood ratio test. The results shown in **Table 4** supported the presence of a random slope in the language ($p = 0.06$), but not in mathematics. Accordingly, the gender gap in mathematics self-concept was stable across classrooms, while the difference between male and female students in language self-concept varied across classrooms. Consequently, in our empirical analysis, we modeled mathematics self-concept using a random intercept model, while allowing a random slope for gender in language self-concept.

TABLE 4 | Results for likelihood ratio tests for random coefficients of gender.

	Language	Mathematics
Female	5.58^{+}	0.09

*Chi-squares based on likelihood ratio test.$^{+}p < 0.10$, *p < 0.05, **p < 0.01, ***p < 0.001.*

In Model 3 in **Tables 2**, **3**, we added students' GSB and the interaction of these beliefs with students' gender. In language, we found no evidence of students' GSB having a significant influence on their self-concept, neither on average nor across male and female students. In mathematics, by contrast, students' GSB had a positive average influence on self-concept and a significantly negative association with gender. Consequently,

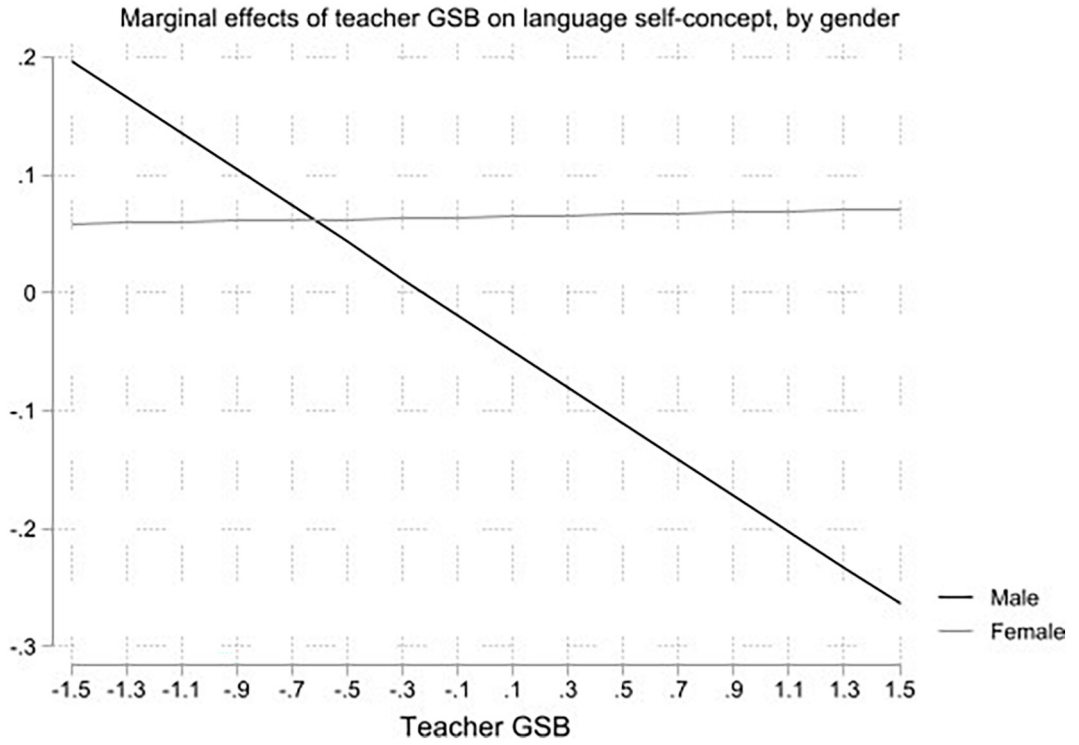


FIGURE 1 | Teacher gender-stereotypical belief (GSB) and predicted language self-concept by gender. Marginal effects of teacher GSB on language self-concept by gender. Based on estimates from Model 6 in **Table 2**.

female students' GSB negatively affected their self-concept in mathematics. In Model 4 in **Tables 2**, **3**, we included peer covariates. We distinguished between the GSB of male and female students to examine if peer influence is gender-specific. Furthermore, we added information on mean peer achievement as well as the gender-achievement pattern in the classroom to investigate how counter-stereotypical achievement patterns influenced students' self-concept in language and mathematics. Mean peer achievement in the classroom had a positive influence on students' self-concept in both language and mathematics. There was no significant impact of peer covariates or their interaction with gender in language. In mathematics, however, we found that the better the mathematics achievement of female peers relative to that of male peers, the lower the mathematics self-concept of girls. In Model 5 in **Tables 2**, **3**, we examined the influence of teachers' GSB and the interaction of these beliefs with students' gender. We found that teachers' stereotypical beliefs influenced students' self-concept in language, but not in mathematics. Accordingly, boys' language self-concept was lower when their teachers endorsed gender stereotypes concerning language ability. Model 6 in **Tables 2**, **3** represented the full model including all student, peer, and teacher covariates and their interactions with gender. Results from this model showed that the relationship between gender and self-concept in language was moderated by the teacher's GSB. Meanwhile, the relationship between gender and self-concept in mathematics was moderated by students' GSB and by counter-stereotypical achievement patterns. These moderation effects remained when taking into account all other covariates and interactions. Consequently, as illustrated in **Figure 1**, our empirical analysis showed that the gender gap in students' self-concept in language increased with teachers' GSB because there was a negative effect on male students.

The gender gap in mathematics was unaffected by teachers' GSB. As illustrated in **Figure 2**, in addition to the GSB and relative performance of female peers, gender differences in mathematics self-concept were mostly affected by students' own GSB. Stronger GSB correlated with higher mathematics self-concept among boys and lower mathematics self-concept among girls, increasing the gender gap.

In addition to students' own GSB, the relative achievement of female/male students also moderated the relationship between gender and self-concept in mathematics. **Figure 3** illustrates that female students on average had lower mathematics self-concept in classrooms with "over-performing" female peers.

Accordingly, being in a counter-stereotypical classroom in terms of stereotypical mathematics achievement did not benefit girls' self-concept in mathematics. On the contrary, the mathematics self-concept of girls was lower when their female peers in the classroom outperformed their male peers in mathematics.

In summary, our empirical analysis showed that 9% of the variation in students' self-concept in language and mathematics

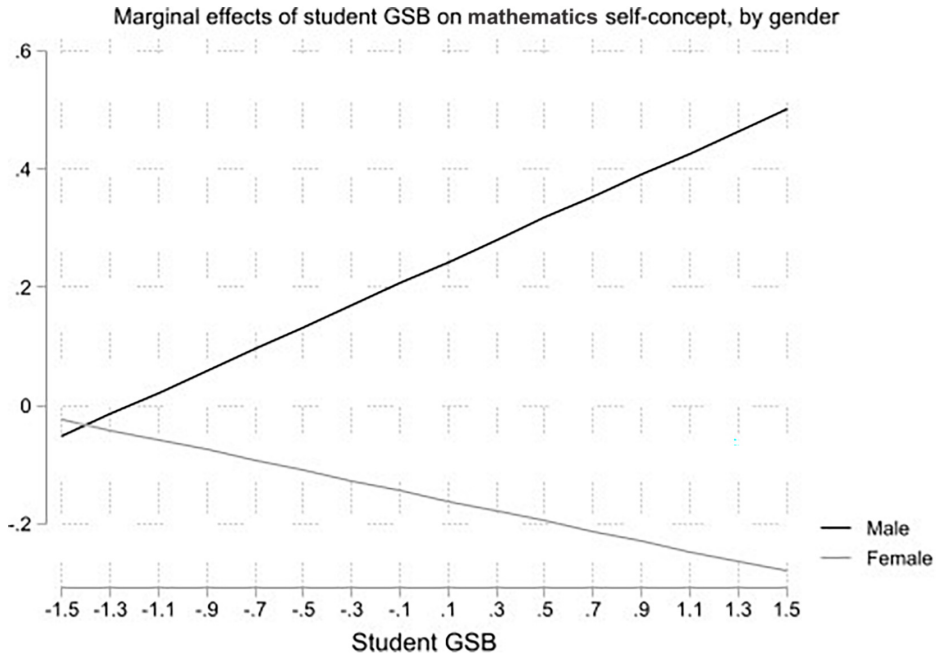


FIGURE 2 | Students' GSB and predicted mathematics self-concept by gender. Marginal effects of teacher GSB on mathematics self-concept by gender. Based on estimates from Model 6 in **Table 3**.

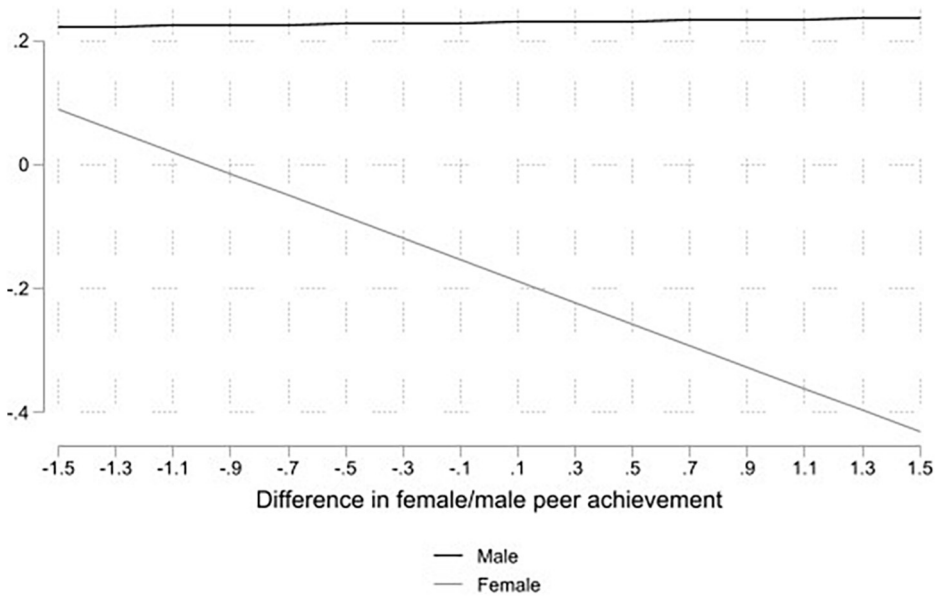


FIGURE 3 | Difference in female/male peer achievement and predicted mathematics self-concept by gender.

was due to variations at the classroom level. The ICC was reduced to approximately 6% in language and approximately 2% in mathematics when introducing school fixed effects to the model. Including contextual covariates further reduced the ICC in both subjects; in the full model, the ICC was close to 0 in mathematics and approximately 3% in language. Furthermore, there was a significant gender gap in students' self-concept in mathematics, but not in language. Gender differences in students' self-concept were unaffected by gender-stereotypical beliefs and achievement patterns since the influence of gender in both subjects was approximately equal across models. Yet, the influence of gender on self-concept varied across classrooms in language, with almost half the variation attributed to interactions between GSB and gender. The random slope of gender in language

was reduced by approximately 28% when introducing teacher GSB. Finally, we found evidence of interaction effects between students' gender and GSB, although the specific mechanism differed across subjects. While teachers' GSB were important for the gender gap in students' self-concept in language, they did not have an impact on gender differences in students' mathematics self-concept. The more teachers endorsed the stereotype that language is for girls, the lower the language self-concept of boys net of actual achievement. By contrast, the influence of students' own gender-stereotypical beliefs differed between male and female students in mathematics. The more girls endorsed the stereotype that mathematics is for boys, the lower their assessment of their mathematical ability. Moreover, girls' self-concept was negatively affected by counter-stereotypical achievement in mathematics; i.e., when female peers in the classroom outperformed male peers.

DISCUSSION

This paper set out to investigate gender differences in students' competence beliefs in language and mathematics and the role of gender beliefs and gender-achievement patterns in the social context of classrooms. We analyzed combined survey and register data from Denmark, which had the advantage that it included measures of GSB at the student, (female/male) peer, and teacher levels, as well as detailed information on student and peer achievement. Furthermore, we only analyzed within-school variation, thus, taking into account selection processes in schools. Our analysis pointed to three main findings. First, consistent with our hypothesis, we found a gender gap in mathematics self-concept favoring boys, even among boys and girls with the same level of performance. This gender gap was quite large (0.39 SD) and did not vary across classrooms. By contrast, the gender gap in language varied across classrooms (p = 0.06). Second, we found that GSB influenced gender gaps in self-concept in both language and mathematics. An increase in GSB was generally associated with a larger gender gap, driven by a decrease in self-concept among students of the negatively stereotyped gender. Accordingly, GSB resulted in a lower self-concept among girls in mathematics and boys in language. Meanwhile, although the general mechanism of GSB was consistent across subject domains, how they operate differed. In language, gender differences in self-concept (which varied across classrooms) were driven by teachers' GSB. As teachers' GSB increased, language self-concept among boys decreased. By contrast, boys' language self-concept was unaffected by their own and male/female peers' GSB. In mathematics, gender differences in self-concept (which did not vary across classrooms) reflected students' own GSB. Accordingly, the greater the extent to which individual (female) students endorsed the stereotype that math is for boys, the lower their mathematical self-concept. Third, counter-stereotypical achievement patterns in mathematics classrooms hurt girls' self-concept. Accordingly, girls did not benefit from being surrounded by female peers that outperformed male peers. Instead, girls in such counter-stereotypical classrooms in terms of mathematics achievement assessed their competence in mathematics at a lower level than girls in classrooms characterized by more gender-stereotypical achievement patterns. This negative influence on achievement among female peers was likely due to a social comparison effect where girls compared themselves to their female classroom peers.

Theoretical and Practical Implications

While our findings support previous empirical and theoretical understandings of the importance of socializing agents (Eccles, 1994) and social contexts (Crosnoe et al., 2008; Legewie and DiPrete, 2014; Riegle-Crumb and Morton, 2017; Salikutluk and Heyne, 2017; Raabe et al., 2019) for the development of competence beliefs, we have also highlighted how these processes differ according to gender and subject domains. Until now, most research has focused on one specific subject domain (such as language *or* mathematics) and gender beliefs and stereotypes among students, peers, *or* teachers (Retelsdorf et al., 2015; Salikutluk and Heyne, 2017; Alan et al., 2018; Muntoni and Retelsdorf, 2018; Heyder et al., 2019; Muntoni et al., 2021). Our study contributes to a greater understanding of the interplay between gender-stereotypical beliefs and achievement patterns and students' self-concept in gender-stereotypical subjects by analyzing very detailed social contexts in classrooms. Accordingly, this paper is to our knowledge the first to consider the multifaceted nature of GSB in schools. While many of our results support what we know from previous studies concerning gender differences in self-concept and the role of GSB, our study adds to this body of research by providing evidence of the complex nature of gender beliefs in schools and their consequences on student outcomes. Most importantly, our findings suggest that how GSB influences students' self-concept are not necessarily homogeneous across gender and subject domains. While students' GSB was associated with students' self-concept in both language and mathematics, the gender gap in language self-concept differed across classrooms within schools. Accordingly, the social context seemed to play a different role in language classrooms than in mathematics classrooms because gender differences in self-concept were related to differences across classrooms. Combined with the finding that language self-concept among boys was most strongly influenced by teacher GSB, this finding suggests that the gender gap in the language is perhaps more malleable. If the language self-concept of male students varies across different social contexts and is influenced by the teacher, then interventions to counteract GSB among teachers might be able to reduce the gender gap in language self-concept.

In contrast to the male disadvantage in language, the gender gap in mathematics was very large and constant across classrooms. This fits well with the finding that girls' mathematics self-concept reflected their own GSB net of their actual mathematics achievement. Accordingly, the female disadvantage in mathematics reflected strong internalized beliefs

about gender and mathematics among the girls themselves. This result may be interpreted in light of cross-cultural research on gender inequality, which has argued that while (vertical) gender inequalities tend to decline in affluent Western democracies, self-expressive value systems, particularly in highly egalitarian countries, still endorse the idea that the genders are innately and fundamentally "equal but different," and therefore continue to encourage the development and enactment of culturally masculine or feminine affinities (Charles and Bradley, 2009). Under these "post-materialist" gender regimes, gender segregation retains legitimacy because it can be understood as the result of free choices by equal, yet innately different, men and women (Thébaud and Charles, 2018). In this context, Denmark can be thought of as a prime example and girls' (boys') strong internalization of GSB in mathematics (language) might be understood as a way of expressing "gendered selves" through cultural gender beliefs. In addition to suggesting a very different mechanism in play than in the case of language, the role of female students' own GSB may also point to a much more stubborn problem. While teachers' GSB about students may be altered through targeted interventions, female students' own internalized perceptions concerning the mathematical ability of girls are most likely harder to change. This is particularly true if such internalized perceptions are intertwined with "natural" and free expressions of gender identity (Cech, 2013) and if girls have a more fixed mindset about math ability with consequences for their competence beliefs (Heyder et al., 2021).

Finally, girls' mathematics self-concept was also influenced by gender-stereotypical achievement patterns of female peers. Our results provided evidence of a social comparison effect, whereby girls evaluated themselves more harshly when surrounded by high-achieving female peers. This result is partly in line with previous research, which has suggested that classmates in general (Salikutluk and Heyne, 2017) and the gender stereotypes of classmates specifically (Muntoni et al., 2021) play an important role in students' educational outcomes. However, we only found this to hold for girls. Consequently, girls may be more prone to the influence of female peers, as has also been suggested by research on gender differences in STEM (Riegle-Crumb et al., 2006; Raabe et al., 2019), which has particularly stressed that intra-gender social comparison effects are stronger than inter-gender effects (Thijs et al., 2010).

In summary, researchers, politicians, and schools should bear in mind that GSB among students, peers, and teachers has a significant and extensive influence on students' competence beliefs and that gender gaps in self-concept in language and mathematics are important for later gender segregation and inequality. Yet, gender beliefs operate in very different ways (Correll, 2001) and need careful consideration and attention, both at the individual and classroom level. Furthermore, our study points to an important distinction between internalized GSB and the GSB of others, which calls for a context-sensitive approach to the consequences for students and the development of possible interventions.

Limitations and Future Directions

The results from this study should be interpreted in light of several limitations. First, the survey data consisted of a non-random sample of schools. During the data collection process, random samples of schools were invited to participate; however, only a fraction of the schools agreed to participate, which meant that schools to some extent self-selected into our sample (although schools could only choose to participate if they had been invited). Analyses of the representativeness of the sample revealed only minor differences between the full population and the sample in terms of parental income and education, as well as 6th-grade test scores (about 0.1 SD). No differences were found in terms of 9th-grade test scores or school size. In other words, the sample was not perfectly representative, but discrepancies in observed characteristics were small (Smith et al., 2021). Furthermore, our original sample of teachers, particularly those teaching mathematics, was significantly reduced due to non-response. We chose not to impute missing data since we did not have good imputation variables at the teacher level. If, for instance, teachers decided to participate in the survey based on certain unobserved characteristics, our results may be biased.

Second, our study aimed to investigate the role of school contexts in the generation of gender disparities in educational outcomes. Yet, we do not know if such disparities – in this case in self-concept – are consequential for later outcomes. An important task for future research is to examine whether early gender disparities translate into later gender differences and inequality in educational behavior and pathways, both within and across gender.

Third, while our explicit measure of GSB was straightforward to implement, there may have been a social desirability bias. Accordingly, we might have underestimated GSB in our sample due to respondents not providing honest answers [deliberately or because of cognitive constraints regarding introspective access to gender perceptions (Wenz et al., 2016)]. Future research should investigate GSB using techniques that have been shown to reduce social desirability bias, such as factorial surveys investigating implicit gender beliefs and their causal mechanisms. Furthermore, future studies could focus on collecting data capable of distinguishing between individual GSB and the perceived GSB of others to explore the internalization of GSB from the social context.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for

participation was not provided by the participants' legal guardians/next of kin in accordance with the local legislation and institutional requirements.

AUTHOR CONTRIBUTIONS

IA and ES contributed to conception and design of the study, and organized the database. IA performed the statistical analysis and wrote the first draft of the manuscript. ES wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the VELUX Foundation, Denmark (Grant Number 00017032).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2022.840618/full#supplementary-material

REFERENCES

Alan, S., Ertac, S., and Mumcu, I. (2018). Gender stereotypes in the classroom and effects on achievement. *Rev. Econ. Statistics* 100, 876–890. doi: 10.1162/rest

Ammermueller, A., and Pischke, J.-S. (2009). Peer effects in European primary schools: evidence from the progress in international reading literacy study. *J. Labor Econ.* 27, 315–348.

Andersen, I. G., and Reimer, D. (2019). Same-gender teacher assignment, instructional strategies, and student achievement: new evidence on the mechanisms generating same-gender teacher effects. *Res. Soc. Stratification Mobility* 62:100406.

Aspenlieder, L., Buchanan, C. M., McDougall, P., and Sippola, L. K. (2009). Gender nonconformity and peer victimization in pre- and early adolescence. *Int. J. Dev. Sci.* 3, 3–16. doi: 10.3233/DEV-2009-3103

Birkelund, J. F., Karlson, K. B., and Reimer, D. (2021). Upper secondary school tracking, labour market outcomes and intergenerational inequality in Denmark. *Long. Life Course Stud.* 12, 279–298.

Bong, M., and Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: how different are they really? *Educ. Psychol. Rev.* 15, 1–40.

Breda, T., Jouini, E., Napp, C., and Thebault, G. (2020). Gender stereotypes can explain the gender-equality paradox. *Proc. Natl. Acad. Sci. U S A.* 117, 31063–31069. doi: 10.1073/pnas.2008704117

Cech, E. A. (2013). The self-expressive edge of occupational sex segregation. *Am. J. Sociol.* 119, 747–789.

Charles, M., and Bradley, K. (2009). Indulging our gendered selves? sex segregation by field of study in 44 countries. *Am. J. Sociol.* 114, 924–976. doi: 10.1086/595942

Correll, S. J. (2001). Gender and the career choice process: the role of biased self-assessments. *Am. J. Sociol.* 10, 1691–1730. doi: 10.1086/321299

Correll, S. J. (2004). Constraints into preferences: gender, status, and emerging career aspirations. *Am. Sociol. Rev.* 69, 93–113.

Cotter, D., Hermsen, J. M., and Vanneman, R. (2011). The end of the gender revolution? gender role attitudes from 1977 to 2008. *Am. J. Sociol.* 117, 259–289. doi: 10.1086/658853

Crosnoe, R., Riegle-Crumb, C., Field, S., Frank, K., and Muller, C. (2008). Peer group contexts of girls' and boys' academic experiences. *Child Dev.* 79, 139–155. doi: 10.1111/j.1467-8624.2007.01116.x

Davis, S. N. (2007). Gender ideology construction from adolescence to young adulthood. *Soc. Sci. Res.* 36, 1021–1041.

DiPrete, T. A., and Buchmann, C. (2013). *The Rise of Women: The Growing Gender Gap in Education and what it Means for American Schools*. New York, NY: Russell Sage Foundation.

Durik, A. M., Vida, M., and Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: a developmental analysis. *J. Educ. Psychol.* 98:382.

Eccles, J. S. (1994). Understanding women's educational and occupational choices: applying the Eccles et al. model of achievement-related choices. *Psychol. Women Quarterly* 18, 585–609.

Eccles, J. S. (2011). Gendered educational and occupational choices: applying the Eccles et al. model of achievement-related choices. *Int. J. Behav. Dev.* 35, 195–201. doi: 10.1177/0165025411398185

Eccles, J. S., and Wigfield, A. (1995). In the mind of the actor: the structure of adolescents' achievement task values and expectancy-related beliefs. *Personal. Soc. Psychol. Bull.* 21, 215–225.

Ellemers, N. (2018). Gender stereotypes. *Annu. Rev. Psychol.* 69, 275–298.

Else-Quest, N. M., Hyde, J. S., and Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychol. Bull.* 136, 103–127. doi: 10.1037/a0018053

England, P. (2010). The gender revolution: uneven and stalled. *Gender Soc.* 24, 149–166. doi: 10.1111/cars.12076

Evans, A. B., Copping, K. E., Rowley, S. J., and Kurtz-Costes, B. (2011). Academic self-concept in black adolescents: do race and gender stereotypes matter? *Self Identity* 10, 263–277. doi: 10.1080/15298868.2010.485358

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., and Yuan, Q. (2018). The big-fish-little-pond effect on academic self-concept: a meta-analysis. *Front. Psychol.* 9:1569. doi: 10.3389/fpsyg.2018.01569

Flore, P. C., and Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? a meta-analysis. *J. School Psychol.* 53, 25–44. doi: 10.1016/j.jsp.2014.10.002

Goldman, A. D., and Penner, A. M. (2016). Exploring international gender differences in mathematics self-concept. *Int. J. Adolescence Youth* 21, 403–418. doi: 10.1080/02673843.2013.847850

Heyder, A., and Kessels, U. (2013). Is school feminine? implicit gender stereotyping of school as a predictor of academic achievement. *Sex Roles* 69, 605–617. doi: 10.1007/s11199-013-0309-9

Heyder, A., Kessels, U., and Steinmayr, R. (2017). Explaining academic-track boys' underachievement in language grades: not a lack of aptitude but students' motivational beliefs and parents' perceptions? *Br. J. Educ. Psychol.* 87, 205–223. doi: 10.1111/bjep.12145

Heyder, A., Steinmayr, R., and Kessels, U. (2019). Do teachers' beliefs about math aptitude and brilliance explain gender differences in children's math ability self-concept? *Front. Educ.* 4:34. doi: 10.3389/feduc.2019.00034

Heyder, A., Weidinger, A. F., and Steinmayr, R. (2021). Only a burden for females in math? gender and domain differences in the relation between adolescents' fixed mindsets and motivation. *J. Youth Adolescence* 50, 177–188. doi: 10.1007/s10964-020-01345-4

Hilton, J. L., and Von Hippel, W. (1996). Stereotypes. *Annu. Rev. Psychol.* 47, 237–271.

Holder, K., and Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: a new look from a shifting standards perspective. *Soc. Psychol. Educ.* 20, 471–490.

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., and Vila-Abad, E. (2010). Polychoric versus pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Qual. Quantity* 44, 153–166.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender similarities characterize math performance. *Science* 321, 494–495. doi: 10.1126/science.1160364

Ireson, J., and Hallam, S. (2009). Academic self-concepts in adolescence: relations with achievement and ability grouping in schools. *Learn. Instruct.* 19, 201–213.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., and Wigfield, A. (2002). Changes in children's self-competence and values: gender and domain differences across grades one through twelve. *Child Dev.* 73, 509–527. doi: 10.1111/1467-8624.00421

Jansen, M., Schroeders, U., and Lüdtke, O. (2014). Academic self-concept in science: multidimensionality, relations to achievement measures, and gender differences. *Learn. Individual Differ.* 30, 11–21.

Lamb, M. E., Easterbrooks, A., and Holden, G. W. (1980). Reinforcement and punishment among preschoolers: characteristics, effects, and correlates. *Child Dev.* 51, 1230–1236.

Leder, G. C., and Forgasz, H. J. (2002). *Two New Instruments To Probe Attitudes about Gender and Mathematics*. Bundoora VIC: La Trobe University.

Legewie, J., and DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *Am. Sociol. Rev.* 77, 463–485.

Legewie, J., and DiPrete, T. A. (2014). The high school environment and the gender gap in science and engineering. *Sociol. Educ.* 87, 259–280. doi: 10.1177/0038040714547770

Loyalka, P., Zakharov, A., and Kuzmina, Y. (2018). Catching the big fish in the little pond effect: evidence from 33 countries and regions. *Comp. Educ. Rev.* 62, 542–564.

Mann, A., Legewie, J., and DiPrete, T. A. (2015). The role of school performance in narrowing gender gaps in the formation of STEM aspirations: a cross-national study. *Front. Psychol.* 6:171. doi: 10.3389/fpsyg.2015.00171

Marsh,H. W. (1987). The big-fish-little-pond effect on academic self-concept. *J. Educ. Psychol.* 79:280.

Marsh, H. W., Craven, R. G., and Martin, A. J. (2006). What is the nature of self-esteem? unidimensional and multidimensional perspectives. *Self-Esteem: Issues Answers* 1, 16–25. doi: 10.1186/s12913-016-1423-5

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., et al. (2019). The murky distinction between self-concept and self-efficacy: beware of lurking jingle-jangle fallacies. *J. Educ. Psychol.* 111:331.

Martinot, D., Bagès, C., and Désert, M. (2012). French children's awareness of gender stereotypes about mathematics and reading: when girls improve their reputation in math. *Sex Roles* 66, 210–219. doi: 10.1007/s11199-011-0032-3

McNally, S. (2020). *Gender Differences in Tertiary Education: What Explains STEM Participation?*. London: Centre for Economic Performance. CEP Discussion Paper No. 1721.

McNeish, D., and Wentzel, K. R. (2017). Accommodating small sample sizes in three-level models when the third level is incidental. *Multivariate Behav. Res.* 52, 200–215. doi: 10.1080/00273171.2016.1262236

Mejía-Rodríguez, A. M., Luyten, H., and Meelissen, M. R. (2021). Gender differences in mathematics self-concept across the world: an exploration of student and parent data of TIMSS 2015. *Int. J. Sci. Mathematics Educ.* 19, 1229–1250.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., and Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Muntoni, F., and Retelsdorf, J. (2018). Gender-specific teacher expectations in reading—the role of teachers' gender stereotypes. *Contemporary Educ. Psychol.* 54, 212–220. doi: 10.1016/j.cedpsych.2018.06.012

Muntoni, F., Wagner, J., and Retelsdorf, J. (2021). Beware of stereotypes: are classmates' stereotypes associated with students' reading outcomes? *Child Dev.* 92, 189–204. doi: 10.1111/cdev.13359

Neuschmidt, O., Barth, J., and Hastedt, D. (2008). Trends in gender differences in mathematics and science (TIMSS 1995-2003). *Stud. Educ. Eval.* 34, 56–72. doi: 10.1016/j.stueduc.2008.04.002

Nowicki, E. A., and Lopata, J. (2017). Children's implicit and explicit gender stereotypes about mathematics and reading ability. *Soc. Psychol. Educ.* 20, 329–345. doi: 10.1007/s11218-015-9313-y

OECD (2013). *Results: Ready to Learn (Volume III): Students' Engagement, Drive and Self-Beliefs*. Paris: OECD Publishing.

OECD (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. Paris: OECD.

Pampel, F. (2011). Cohort changes in the socio-demographic determinants of gender egalitarianism. *Soc. Forces* 89, 961–982. doi: 10.1353/sof.2011.0011

Pansu, P., Régner, I., Max, S., Colé, P., Nezlek, J. B., and Huguet, P. (2016). A burden for the boys: evidence of stereotype threat in boys' reading performance. *J. Exp. Soc. Psychol.* 65, 26–30. doi: 10.1016/j.jesp.2016.02.008

Penner, A. M. (2008). Gender differences in extreme mathematical achievement: an international perspective on biological and social factors. *Am. J. Sociol.* 114, S138–S170. doi: 10.1086/589252

Plante, I., De la Sablonnière, R., Aronson, J. M., and Théorêt, M. (2013). Gender stereotype endorsement and achievement-related outcomes: the role of competence beliefs and task values. *Contemporary Educ. Psychol.* 38, 225–235. doi: 10.1016/j.cedpsych.2013.03.004

Plante, I., Théorêt, M., and Favreau, O. E. (2009). Student gender stereotypes: contrasting the perceived maleness and femaleness of mathematics and language. *Educ. Psychol.* 29, 385–405. doi: 10.1080/01443410902971500

Plieninger, H., and Dickhäuser, O. (2015). The female fish is more responsive: gender moderates the BFLPE in the domain of science. *Educ. Psychol.* 35, 213–227.

Raabe, I. J., Boda, Z., and Stadtfeld, C. (2019). The social pipeline: how friend influence and peer exposure widen the STEM gender gap. *Sociol. Educ.* 92, 105–123.

Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLoS One* 7:e39904. doi: 10.1371/journal.pone.0039904

Reilly, D., Neumann, D. L., and Andrews, G. (2019). Gender differences in reading and writing achievement: evidence from the National Assessment of Educational Progress (NAEP). *Am. Psychol.* 74:445. doi: 10.1037/amp0000356

Retelsdorf, J., Schwartz, K., and Asbrock, F. (2015). "Michael can't read!" teachers' gender stereotypes and boys' reading self-concept. *J. Educ. Psychol.* 107, 186–194. doi: 10.1037/a0037107

Ridgeway, C. L. (2009). Framed before we know it: how gender shapes social relations. *Gender Soc.* 23, 145–160.

Ridgeway, C. L., and Correll, S. J. (2004). Unpacking the gender system: a theoretical perspective on gender beliefs and social relations. *Gender Soc.* 18, 510–531. doi: 10.1177/1557988320927285

Riegle-Crumb, C., and Morton, K. (2017). Gendered expectations: examining how peers shape female students' intent to pursue STEM fields. *Front. Psychol.* 8:329. doi: 10.3389/fpsyg.2017.00329

Riegle-Crumb, C., and Peng, M. (2021). Examining high school students' gendered beliefs about math: predictors and implications for choice of STEM college majors. *Sociol. Educ.* 94:003804072110147.

Riegle-Crumb, C., Farkas, G., and Muller, C. (2006). The role of gender and friendship in advanced course taking. *Sociol. Educ.* 79, 206–228. doi: 10.1177/003804070607900302

Riegle-Crumb, C., King, B., Grodsky, E., and Muller, C. (2012). The more things change, the more they stay the same? prior achievement fails to explain gender inequality in entry into STEM college majors over time. *Am. Educ. Res. J.* 49, 1048–1073. doi: 10.3102/0002831211435229

Riegle-Crumb, C., Kyte, S. B., and Morton, K. (2018). "Gender and racial/ethnic differences in educational outcomes: examining patterns, explanations, and new directions for research bt - handbook of the sociology of education in the 21st Century," in , *Handbook of the Sociology of Education in the 21st Century*, ed. B. Schneider (Berlin: Springer International Publishing).

Risman, B. J. (2004). Gender as a social structure: theory wrestling with activism. *Gender Soc.* 18, 429–450.

Rosenqvist, E. (2018). Two functions of peer influence on upper-secondary education application behavior. *Sociol. Educ.* 91, 72–89.

Salikutluk, Z., and Heyne, S. (2017). Do gender roles and norms affect performance in maths? the impact of adolescents' and their peers' gender conceptions on maths grades. *Eur. Sociol. Rev.* 33, 368–381.

Seaton, M., Marsh, H. W., and Craven, R. G. (2010). Big-fish-little-pond effect: generalizability and moderation—two sides of the same coin. *Am. Educ. Res. J.* 47, 390–433.

Sikora, J., and Pokropek, A. (2012). Gender segregation of adolescent science career plans in 50 countries. *Sci. Educ.* 96, 234–264. doi: 10.1002/sce.20479

Skaalvik, S., and Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles* 50, 241–252.

Smith, E., Reimer, D., Andersen, I. G., and Sortkær, B. (2021). *Exploring School Culture: Technical Report for Data Collection. Danish School of Education*. Available online at: https://projekter.au.dk/fileadmin/projekter/Exploring_School_Culture/Ebog_-_ESCU_technical_report.pdf (accessed December 15, 2021).

Steele, C. M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *Am. Psychol.* 52:613. doi: 10.1037//0003-066x.52.6.613

Thébaud, S., and Charles, M. (2018). Segregation, stereotypes, and STEM. *Soc. Sci.* 7:111.

Thijs, J., Verkuyten, M., and Helmond, P. (2010). A further examination of the big-fish-little-pond effect: perceived position in class, class size, and gender comparisons. *Sociol. Educ.* 83, 333–345.

UNESCO (2017). *Cracking the Code: Girls' and Women's Education in Science, Technology, Engineering and Mathematics (STEM)*. Paris: UNESCO.

Vincent-Lancrin, S. (2008). "The reversal of gender inequalities in higher education: an on-going trend," in *Higher Education to 2030: Demography*, Vol. 1 (Paris: OECD Publishing). doi: 10.1787/9789264040663-11-en

Voyer, D., and Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychol. Bull.* 140:1174. doi: 10.1037/a0036620

Wenz, S. E., Olczyk, M., and Lorenz, G. (2016). *Measuring Teachers' Stereotypes in the NEPS*. Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. NEPS Survey Paper, 3.

Wigfield, A., and Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educ. Psychol.* 25, 68–81.

Witt, S. D. (2000). The influence of peers on children's socialization to gender roles. *Early Child Dev. Care* 162, 1–7.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 19 May 2022
doi: 10.3389/fpsyg.2022.856758

Benefits of Psychological Androgyny in Adolescence: The Role of Gender Role Self-Concept in School-Related Well-Being

Selma Korlat[1]†, Julia Holzer[1†], Marie-Therese Schultes[2†], Sarah Buerger[1†], Barbara Schober[1†], Christiane Spiel[1†] and Marlene Kollmayer[1†]*

[1]Department for Psychology of Development and Education, Faculty of Psychology, University of Vienna, Vienna, Austria, [2]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:
Joyce Endendijk,
Utrecht University, Netherlands

Reviewed by:
Carmen Martínez,
University of Murcia, Spain
Matthew Nielson,
University of Michigan, United States

***Correspondence:**
Selma Korlat
selma.korlat@univie.ac.at

†ORCID:
Selma Korlat
orcid.org/0000-0001-5719-5222
Julia Holzer
orcid.org/0000-0002-0029-3291
Marie-Therese Schultes
orcid.org/0000-0002-1593-3247
Sarah Bürger
orcid.org/0000-0002-4645-9701
Barbara Schober
orcid.org/0000-0002-5754-8865
Christiane Spiel
orcid.org/0000-0002-7082-4879
Marlene Kollmayer
orcid.org/0000-0001-8239-5330

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: 17 January 2022
Accepted: 06 April 2022
Published: 19 May 2022

Citation:
Korlat S, Holzer J, Schultes M-T, Buerger S, Schober B, Spiel C and Kollmayer M (2022) Benefits of Psychological Androgyny in Adolescence: The Role of Gender Role Self-Concept in School-Related Well-Being. Front. Psychol. 13:856758. doi: 10.3389/fpsyg.2022.856758

It has been repeatedly shown that the extent to which individuals adopt stereotypically masculine and feminine traits in their self-concept impacts their health and well-being. This is especially important in adolescence, when developmental changes and social pressures to conform to stereotypical gender roles can affect psychological functioning. However, previous studies investigating relationship between gender role self-concept and well-being in adolescents focused mostly on general well-being rather than well-being in specific contexts. Given that school is one of the most important contexts for adolescents' development and well-being, the aim of this study was to investigate differences between adolescents with different gender role self-concepts (masculine, feminine, androgynous and undifferentiated) in *school-related well-being*. In line with the new conceptualization of well-being uniting hedonic (pleasure attainment and pain avoidance) and eudemonic (self-actualization and having meaningful purpose in one's life) approaches, the present study used a measure of school-related well-being encompassing five domains suggested in the EPOCH (Engagement, Perseverance, Optimism, Connectedness and Happiness) model as well as a superordinate well-being factor. A total of 999 Austrian adolescents (52.2% girls, M_{age} = 13.79, SD_{age} = 1.53) answered inventories assessing adolescents' gender role self-concept (GRI-JUG) and school-related well-being (EPOCH-G-S). The results supported the androgyny model of well-being, showing clear advantages of having both positive masculine and feminine qualities in one's self-concept for optimal levels of school-related well-being. In addition, our results indicated the strong importance of femininity in adolescence and the school context. Theoretical and practical implications are discussed.

Keywords: well-being, hedonic, eudemonic, school, gender, androgyny, adolescence

INTRODUCTION

Adolescence is seen as a key stage of development characterized by profound changes in the biological and psychosocial domains (Choudhury et al., 2006). In this period, individuals should develop the capabilities required to lead a happy, healthy, and productive life (Laski, 2015). According to gender intensification theory, it is during this same time that girls and

boys develop increasingly differentiated gender role identities due to increased pressure to conform to stereotypical gender roles (Hill and Lynch, 1983). These two developmental processes are interrelated, as it has been repeatedly shown that differences in the extent to which individuals adopt stereotypically masculine and feminine traits in their self-concept impact their psychological adjustment and well-being (Abele, 2014; Wang, 2016; Martínez-Marín and Martínez, 2019; Matud et al., 2019). Likewise, high levels of well-being in adolescence enable young people to deal with developmental tasks and promote a healthy transition to adulthood (Pyhältö et al., 2010). It has been shown that adolescents' well-being is related to their academic functioning (Lewis et al., 2011), academic achievement (Berger et al., 2011), and is a protective factor for health in general (Carver et al., 2010). Studies have shown that adolescents with high levels of well-being are more resilient (Gilman and Huebner, 2006; Antaramian et al., 2010), have fewer depressive and anxiety symptoms, higher self-esteem, higher self-efficacy and higher adaptation (McKnight et al., 2002; Antaramian et al., 2010). These adolescents show enhanced mental and physical health outcomes and higher general life satisfaction (Tian et al., 2014).

Well-being in adolescence is integrally shaped by the everyday contexts in which adolescents grow and develop (Žukauskienė, 2014). As the place where adolescents spend almost one-third of their lives, school is one of the most important contexts within which adolescents' development, including their well-being and gender socialization, unfolds (Eccles, 2004). It has been shown that adolescents' experiences and relationships at school have an important impact on their perceived quality experiences and relationships in school of life (Jourdan et al., 2008) and likely have important implications for their lifelong development (Park, 2004; Eccles and Roeser, 2011). Given the developmental milestones related to gender identity and well-being in adolescence, as well as the importance of school in this period of life, this study aims to understand the role of gender role self-concept in well-being in the school context.

Gender Role Self-Concept and Well-Being

Socialization pressure may lead adolescents to internalize societal gender role expectations as part of their self-concept (Bem, 1981; Hill and Lynch, 1983; Klaczynski et al., 2020). Gender role self-concept refers to the degree to which persons adopt stereotypically feminine and masculine attributes in their self-descriptions (Wolfram et al., 2009). In most studies investigating gender role self-concept, self-perception of expressive traits (e.g., being kind, gentle, sensitive to others) is used to assess femininity, and self-perception of instrumental traits (e.g., being independent, competitive, strong) is used to assess masculinity (Bem, 1974, 1981; Spence, 1991). While the equating of instrumentality (or agency) with masculinity, and expressiveness (or communion) with femininity has previously been questioned and criticized (e.g., Pedhazur and Tetenbaum, 1979), a more recent principal component analysis confirmed their interchangeability, showing that these concepts can be equated on an operational level (Abele and Wojciszke, 2007). Similarly, previous studies investigating the role of gender role self-concept in well-being labeled these gendered dimensions interchangeably when referring to the same concepts—self-ascribed gender stereotyped traits (see, e.g., Yarnell et al., 2019). Hence, we use and interpret these terms synonymously. Regardless of labels, four types of gender role self-concepts can be formulated based on continuous scores on these dimensions. *Masculine individuals* perceive themselves as high on masculine and low on feminine traits, *feminine individuals* score high on feminine and low on masculine traits, *androgynous persons* rate themselves as high on both sets of traits, and *undifferentiated individuals* view themselves as low on both sets of traits. The interrelations between gender role self-concept and well-being have been widely studied, as it has been noted that individual differences in these two dimensions affect overall functioning and health (e.g., Abele, 2014; Martínez-Marín and Martínez, 2019; Matud et al., 2019). There are three different models explaining the relationship between gender role self-concept and well-being, namely the congruence model, the androgyny model and the masculinity model.

Traditionally, individuals' psychological well-being was thought to be related to their successful adoption of gender-typical behaviors and traits (the congruence model, Markstrom-Adams, 1989; DiDonato and Berenbaum, 2011). This hypothesis has received some support with preadolescent samples. For instance, Carver et al. (2003) found that early adolescents who perceived themselves to be atypical members of their same-sex peer group reported distress over their peer relations. In another study, the same authors confirmed that feeling gender-typical was positively related to adolescents' well-being, whereas feeling pressure to conform to gender stereotypes was found to have a negative influence (Yunger et al., 2004). These authors did not measure gender role self-concept as the self-ascription of gender-typical traits, but focused on a measure of gender identity based on adolescents' feelings of same-gender typicality, which is composed of five different components: membership knowledge, gender typicality, gender contentedness, felt pressure for gender conformity and intergroup bias (Egan and Perry, 2001). Other studies using Egan and Perry's (2001) measure have also found that the more adolescents feel same-gender typical, the greater their self-esteem and the fewer internalizing problems they have (Corby et al., 2007; Menon et al., 2013; Pauletti et al., 2017).

The androgyny model (Bem, 1974; Spence and Helmreich, 1979) posits that psychological well-being is maximized when one has an androgynous gender role self-concept, which encompasses a broad set of attributes and behavioral options that allow for flexible behavior and successful coping with different demands and life situations. Studies have found that women and men whose self-concept includes both masculine-instrumental and feminine-expressive characteristics have greater well-being (e.g., Wang, 2016; Matud et al., 2019). More recent studies using both—a new measure of gender identity (Egan and Perry, 2001; Pauletti et al., 2017) and self-ascribed gender typical attributes (Martínez-Marín and Martínez, 2019), showed benefits of androgyny for well-being, self-esteem and psychological adaptation of adolescents. However, the proposed relationship between androgyny and psychological well-being has been called into question by empirical findings claiming

that it is the masculine component of androgyny that is most associated with both adolescents' and adults' well-being (the masculinity model; Whitley, 1985). Indeed, the vast majority of studies have found masculinity to be associated with subjective well-being and other self-report measures of psychological adjustment (e.g., Wolfram et al., 2009; Abele et al., 2016; Matud et al., 2019). Studies with adolescents have come to more heterogeneous results. In an old study with adolescents aged 11, 13 and 15 in the United States using the Children's Sex Role Inventory (CSRI; Boldizar, 1991), masculinity was linked to lower rates of depressive symptoms, while femininity was not (Priess et al., 2009). Another cross-sectional study using a brief version of the Bem Sex Role Inventory (BSRI) with 12,287 Norwegian adolescents aged 12 to 20 found femininity to be modestly positively correlated with depressed mood, whereas no such correlation was obtained for masculinity (Wichstrøm, 1999). On the other hand, Helgeson and Palladino (2012) found that both masculinity and femininity, measured as agentic and communal traits, were associated with positive relationship and health outcomes among US adolescents, with femininity being a stronger predictor than masculinity. A recent longitudinal study with Chinese children and adolescents aged 6–11 using self-descriptive questionnaires containing instrumental and expressivity traits found that older children's self-esteem was more related to instrumental than expressive traits, whereas younger children's self-esteem was more related to expressivity than instrumentality (Chen et al., 2018). These authors argued that expressive traits and behaviors are relatively more important to younger children's self-esteem due to the prominence of social goals at this age; conversely, instrumental traits and behaviors are relatively more important to older children's self-esteem due to the increasing importance of performance-related goals. In sum, empirical evidence supports both the masculinity and androgyny models of well-being, but indicates the stronger importance of femininity in adolescence compared to adulthood.

School-Related Well-Being

Well-being has also been operationalized in heterogeneous ways in existing studies, ranging from self-esteem measures (e.g., Carver et al., 2003), positive and negative affect and life satisfaction scales (e.g., Buchanan and Bardi, 2015), absence of depression (e.g., Priess et al., 2009) and psychological distress (Helgeson and Palladino, 2012), to various measures of adjustment such as low internalizing problems (e.g., Pauletti et al., 2017). On the whole, most previous measures focused on hedonic well-being, which defines well-being in terms of attaining pleasure and avoiding pain (Kahneman, 1999), and refer to well-being as an outcome. However, in recent years, with the emergence and growth of positive psychology, well-being has been reconceptualized in a way that includes eudemonic aspect as well. This aspect refers to self-actualization and having meaningful purpose in one's life, defining well-being in terms of personal growth experience (Tian et al., 2014). This led to the conceptualization of well-being as consisting of both hedonic and eudemonic dimensions in terms of the full functioning of the person, referring to well-being as a process (Ryan and Deci, 2001). Against this backdrop, Seligman (2011) proposed a five-element model consisting of Positive Emotions, Engagement, Relationships, Meaning, and Accomplishment (PERMA). Applying the PERMA model to adolescents (Kern et al., 2015) led to the development of the EPOCH model of adolescent well-being (Kern et al., 2016), which likewise encompasses five domains: Engagement, Perseverance, Optimism, Connectedness, and Happiness. While optimism and happiness correspond to the hedonic aspect of well-being, other EPOCH dimensions capture eudemonic characteristics. Engagement refers to the capacity to become absorbed in and focused on activities and tasks. Perseverance reflects the ability to keep striving toward one's goals despite encountering obstacles. Optimism is characterized by hopefulness and confidence about the future. Connectedness describes satisfying relationships and friendships, giving and receiving support to and from others. Happiness refers to having a generally positive mood and feeling content with one's life (Kern et al., 2016). Thus, EPOCH covers a wide range of components associated with optimal functioning in adolescence, taking into account both hedonic as well as eudemonic aspects of well-being. Hence, it unites domains that have been studied individually or combined under the umbrella term "well-being" in different constellations.

Another change resulting from positive psychology is the emphasis on youth's optimal well-being in specific contexts (Seligman and Csikszentmihalyi, 2000; Elmore and Huebner, 2010; Long et al., 2012). School is a key life context for school-aged children and adolescents and a place where they spend a great deal of time. As such, school plays an important role in every aspect of youth development—it shapes their identity, intellectual and cognitive growth, social relationships and psychological well-being (Park, 2004; Long and Huebner, 2014; Verhoeven et al., 2019). Moreover, school is a place where adolescents prepare for their future (Eccles and Roeser, 2011). Understanding adolescents' overall functioning and well-being in school is of utmost importance. However, most existing scales for subjective school-related well-being encompassed one cognitive element linked to the individual's life satisfaction and two affective—positive affect and negative affect (see Liu et al., 2016) or used indicators of adolescents' hedonic well-being and functioning in school (e.g., global school satisfaction or achievement; see Kern et al., 2015 or Yang et al., 2018 for further discussion). These indicators are, however, not considered as components of adolescents' psychological well-being that focus on subjective experience (Holzer et al., 2021). Accordingly, previous studies showed well-being to be an important but distinguishable correlate of other indicators of academic functioning (e.g., Howell, 2009; Steinmayr et al., 2018). In line with the new conceptualization of well-being uniting hedonic and eudemonic approaches, in the present study a new measure of school-related well-being has been used: Buerger et al. (2022) have adapted the EPOCH model to the school context, resulting in the EPOCH-School (i.e., EPOCH-S), and its corresponding measure in German language, the EPOCH-German-School (i.e., EPOCH-G-S). Thus, engagement and perseverance according to the EPOCH-S model refer to school tasks and activities. Optimism refers to positive expectations of future academic success and future experiences at school. Connectedness refers

to positive relationships in school in general, whether with peers or teachers, and happiness refers to positive mood in school and satisfaction with school life. By explicitly focusing on eudemonic and hedonic aspects of well-being, the EPOCH-S model distinguishes itself from previous operationalizations of school well-being *via* general school-related emotions or outcomes, reflecting the full variety of adolescents' functioning in the school context. Moreover, the psychological characteristics and processes as indicators of well-being construct in the EPOCH-S model allow to derive specific intervention needs that might promote more global outcomes such as achievement or school satisfaction.

Objectives of the Present Study

This study focuses on the relations between adolescents' gender role self-concept and school-related well-being in terms of the EPOCH-S model. We investigate differences between adolescents with different gender role self-concepts (masculine, feminine, androgynous and undifferentiated) in terms of overall school-related well-being (EPOCH-S) as well as in the individual EPOCH-S dimensions: Engagement, Perseverance, Optimism, Connectedness, and Happiness.

Taking into account both social and performance-related challenges and requirements in school and the fact that androgynous individuals have the broadest repertoire of traits and behaviors (e.g., Bem, 1981; Pauletti et al., 2017), we expect androgynous boys and girls to show the highest levels of overall school-related well-being compared to the other three types of gender role self-concept. Similarly, we expect androgynous boys and girls to have the highest levels of the two hedonic dimensions—optimism and happiness. We have differential hypotheses for the eudemonic dimensions. Engagement and perseverance are instrumental qualities, as they reflect orientations toward achievement-related goals, while connectedness reflects the core value of expressivity—orientation towards others and social-related goals (Abele et al., 2016). Therefore, we expect masculine and androgynous boys and girls to exhibit the highest levels of engagement and perseverance and feminine and androgynous boys and girls to exhibit the highest levels of connectedness. Undifferentiated adolescents are expected to have the lowest levels of both overall school-related well-being as well as all individual dimensions compared to the other three types. We expected our hypotheses to be confirmed for both boys and girls of all ages in our study. In order to investigate potential changes in school-related well-being throughout middle adolescence, age was included as predictor in all analyses.

MATERIALS AND METHODS

Participants and Procedure

The study sample consisted of 999 secondary school students (52.2% girls, $M_{age} = 13.79$, $SD_{age} = 1.53$; age range 12–17) from Vienna, Austria. Data collection took place in January 2020. To recruit participants, 10 secondary schools in Vienna were contacted by e-mail. Seven of these schools agreed to participate and were included in the sample. Students filled out paper-pencil questionnaires in their classrooms, supervised by trained research assistants. Participation in the study was voluntary and parental consent was obtained for participation in the study as well as data usage. The study was approved and supported by the local school board in accordance with Austrian federal law.

Measures

Gender Role Self-Concept

To assess self-perceived femininity and masculinity, positive traits from the Inventory for Measuring Adolescents' Gender Role Self-concept (GRI-JUG) were used (Krahé et al., 2007). Although the original GRI-JUG instrument encompasses negative traits as well, we assessed only positive traits in our study to keep the questionnaire's length reasonable. This is in line with Bem's theorizing (1981) and related studies (e.g., Woodhill and Samuels, 2003), showing that only positive androgyny, that is a balance of positive masculine and positive feminine qualities, is relevant for psychological well-being. Participants were presented with five masculine attributes (humorous, courageous, sporty, companionable, and strong; $\alpha = 0.68$) and five feminine attributes (emotional, romantic, industrious, sympathetic and empathic; $\alpha = 0.66$), and were asked to rate to what extent each attribute is characteristic of them on a 5-point Likert scale ranging from 1 (not at all true) to 5 (completely true). Scores were calculated for masculinity and femininity separately. The median-split procedure adopted by Spence et al. (1975) and Bem (1977) was used to determine the four types of gender role self-concept (see **Table 1**). Participants were classified into a 2 × 2 table according to whether they fell above or below the median score on the masculinity and femininity scales. Scores falling exactly on the median were classified as "high" scores (e.g., Carver et al., 2013). In the present sample, the median masculinity score was 3.8 and the median femininity score was 3.6. All four types significantly differed in both masculinity [$F(3,993) = 681.094$, $p < 0.001$] and femininity [$F(3,993) = 669.293$, $p < 0.001$], with androgynous adolescents scoring highest on both dimensions ($M = 4.36$ for masculinity and $M = 4.12$ for femininity), followed by masculine ($M = 4.19$) and feminine ($M = 3.30$) type on masculinity and feminine ($M = 3.96$) and masculine (M = 3.07) type on femininity. Undifferentiated type scored significantly lowest on both dimensions (M = 3.06 for masculinity and $M = 2.85$ for femininity) compared to other types.

TABLE 1 | Gender role self-concept types.

		Femininity	
		Low	**High**
Masculinity	Low	Undifferentiated	Feminine
	High	Masculine	Androgynous

Low = scores smaller than the median, high = scores higher than the median.

Well-Being

School-related well-being was assessed with the EPOCH-G-S Measure of School-related Adolescent Well-Being (Buerger et al., 2022), a 19-item measure developed for students aged 10–18. The EPOCH-G-S measure of student's well-being in school was validated with results favoring a second-order model with well-being as a second-order factor and the five specific EPOCH-S first-order factors. Invariance analyzes showed scalar invariance, indicating that factor means can be compared between boys and girls as well as between different age groups. Thus, the EPOCH-G-S with its multidimensional structure allows for detecting strengths and weaknesses in students' well-being profiles and intervene on the school, class or individual level. The measures' scales address the five dimensions of the EPOCH-S model: Engagement (four items, e.g., "When I do an activity for school, I enjoy it so much that I lose track of time"), Perseverance (four items, e.g., "When I have started a school task, I finish it"), Optimism (three items, e.g., "I am optimistic about my future at school"), Connectedness (four items, e.g., "When something good happens to me, I have people at school who I like to share the good news with"), and Happiness (four items, e.g., "I feel happy at school"). The measure uses a 5-point response format (1 = not true at all; 5 = completely true). The internal reliability of the EPOCH-G-S was $\alpha = 0.86$. Reliabilities for the EPOCH-G-S subscales were $\alpha = 0.72$ for Engagement, $\alpha = 0.79$ for Perseverance, $\alpha = 0.67$ for Optimism, $\alpha = 0.73$ for Connectedness, and $\alpha = 0.85$ for Happiness.

RESULTS

Overall School-Related Well-Being

In order to examine differences in overall school-related well-being in adolescents, a 4×2 ANCOVA was conducted with gender role self-concept and sex as between-subject factors and age as a covariate. The mean score of all EPOCH-G-S items was the dependent variable. Means and standard deviations for overall school-related well-being by gender role self-concept and sex are presented in **Table 2**.

The results showed a significant effect of age, $F(1, 996) = 27.72$, $p < 0.001$, $\eta^2 p = 0.027$, indicating a negative relationship between age and school-related well-being, $r(998) = -0.147$, $p < 0.001$. There was also a significant main effect of gender role self-concept after controlling for adolescents' age, $F(3, 996) = 60.98$, $p < 0.001$, $\eta^2 p = 0.156$. A Bonferroni *post hoc* test showed that androgynous adolescents ($M = 3.73$, $SD = 0.54$) reported significantly higher overall school-related well-being than masculine ($M = 3.39$, $SD = 0.53$), feminine ($M = 3.42$, $SD = 0.49$) and undifferentiated ($M = 3.16$, $SD = 0.57$) adolescents, all $ps < 0.001$. Undifferentiated adolescents, on the other hand, had significantly lower overall school-related well-being than adolescents with other gender role self-concepts, all $ps < 0.01$. There were no differences between masculine and feminine adolescents in overall school-related well-being, $p > 0.05$. The main effect of sex was also significant, $F(1, 996) = 16.21$, $p < 0.001$, $\eta^2 p = 0.016$, with girls ($M = 3.50$, $SD = 0.53$) reporting higher overall school-related well-being compared to boys ($M = 3.35$,

TABLE 2 | Means and standard deviations for overall school-related well-being and single EPOCH-S dimensions by gender role self-concept and sex.

Gender role self-concept	Sex	*N*	Overall well-being		Engagement		Perseverance		Optimism		Connectedness		Happiness	
			M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Androgynous	Boys	163	3.71^{a}	0.54	3.07^{a}	0.85	3.81^{a}	0.74	3.72^{a}	0.81	4.15^{a}	0.72	3.77^{a}	0.87
	Girls	216	3.78^{a*}	0.54	3.08^{a*}	0.80	3.93^{a*}	0.77	3.57^{a}	0.79	4.34^{a*}	0.68	3.86^{a*}	0.86
Masculine	Boys	128	3.26^{b}	0.53	2.61^{b}	0.85	$3.29^{b,c}$	0.85	3.25^{b}	0.83	3.85^{b}	0.68	3.32^{b}	0.91
	Girls	58	3.54^{b*}	0.53	2.79^{b*}	0.74	$3.67^{b,c*}$	0.83	3.41^{b}	0.88	4.19^{b*}	0.73	3.61^{b*}	0.80
Feminine	Boys	53	3.32^{b}	0.49	2.64^{b}	0.75	3.60^{b}	0.78	3.37^{b}	0.74	3.76^{b}	0.90	3.25^{b}	0.93
	Girls	121	3.48^{b*}	0.49	2.80^{b*}	0.77	3.62^{b*}	0.75	3.32^{b}	0.82	4.15^{b*}	0.77	3.47^{b*}	0.84
Undifferentiated	Boys	133	3.11^{c}	0.59	2.56^{b}	0.89	3.49^{c}	0.84	3.14^{c}	0.88	3.51^{c}	0.83	3.14^{c}	0.89
	Girls	124	3.22^{c*}	0.55	2.76^{b*}	0.80	3.71^{c*}	0.80	2.91^{c}	0.80	3.76^{c*}	0.86	3.18^{c*}	0.84

*Means not sharing the same (a, b, c) superscripts within column are significantly different, with superscripts from a to c indicating scores from highest to lowest. An * indicates significant gender differences for each dimension with * being placed next to the higher mean.*

TABLE 3 | Correlations between EPOCH-S dimensions.

	1	2	3	4	5
Engagement	1	0.553**	0.341**	0.076*	0.349**
Perseverance		1	0.426**	0.141**	0.390**
Connectedness			1	0.274**	0.584**
Optimism				1	0.425**
Happiness					1

*$p < 0.05$; **$p < 0.01$.

$SD = 0.54$). The interaction between gender role self-concept and sex was not significant, $F(3, 996) = 1.58$, $p = 0.192$.

Dimensions of School-Related Well-Being

To explore differences between boys and girls with different gender role self-concepts in the EPOCH-G-S dimensions, a two-way multivariate analysis of covariance (MANCOVA) with gender role self-concept and sex as between-subject factors and age as a covariate was conducted. The mean scores of the five EPOCH-G-S dimensions Engagement, Perseverance, Optimism, Connectedness and Happiness served as dependent variables in the model. Correlation coefficients between EPOCH-G-S dimensions are reported in **Table 3**. The MANCOVA yielded significant multivariate effects for gender role self-concept, $F(3, 996) = 12.86$, $p < 0.001$, $\eta^2p = 0.061$, and sex, $F(1, 996) = 11.76$, $p < 0.001$, $\eta^2p = 0.056$, as well as the covariate age, $F(1, 996) = 14.30$, $p < 0.001$, $\eta^2p = 0.068$. The interaction effect between factors was not significant, $F(3, 996) < 1$, $p = 0.476$. We followed up on the significant multivariate effects with univariate analyses of covariance (ANCOVAs). **Table 2** presents the means and standard deviations for all five EPOCH-G-S dimensions by gender role self-concept and sex.

Engagement

The results showed a significant effect of age, $F(1, 996) = 16.88$, $p < 0.001$, $\eta^2p = 0.017$. There was a negative relationship between age and engagement, $r(998) = -0.127$, $p < 0.001$. There was also a significant main effect of gender role self-concept after controlling for age, $F(3, 996) = 17.46$, $p < 0.001$, $\eta^2p = 0.050$. A Bonferroni *post hoc* test revealed that androgynous adolescents ($M = 3.07$, $SD = 0.82$) reported significantly higher scores on engagement than masculine ($M = 2.69$, $SD = 0.82$), feminine ($M = 2.74$, $SD = 0.76$), and undifferentiated ($M = 2.65$, $SD = 0.85$) adolescents. Differences among the other types were not significant, al $ps > 0.05$. The main effect of sex was also significant after controlling for age, $F(1, 996) = 5.58$, $p < 0.05$, $\eta^2p = 0.006$, with girls ($M = 2.86$, $SD = 0.80$) reporting higher engagement than boys ($M = 2.72$, $SD = 0.88$).

Perseverance

The results showed a significant effect of age, $F(1, 996) = 38.28$, $p < 0.001$, $\eta^2p = 0.037$. The relationship between age and perseverance was negative, $r(998) = -0.181$, $p < 0.001$. The results showed also a significant main effect of gender role self-concept after controlling for age, $F(3, 996) = 29.23$, $p < 0.001$, $\eta^2p = 0.082$. A Bonferroni *post hoc* test revealed that androgynous adolescents ($M = 3.87$, $SD = 0.76$) reported significantly higher scores on perseverance than masculine ($M = 3.46$, $SD = 0.85$), feminine ($M = 3.64$, $SD = 0.76$), and undifferentiated ($M = 3.31$, $SD = 0.81$) adolescents, $p < 0.05$. Undifferentiated adolescents reported lower scores than androgynous and feminine adolescents, $p < 0.05$. Mean differences between the other groups were not significant, $p > 0.05$. The main effect of sex was also significant after controlling for age, $F(1, 996) = 10.40$, $p < 0.01$, $\eta^2p = 0.010$, with girls ($M = 3.66$, $SD = 0.80$) reporting higher perseverance than boys ($M = 3.48$, $SD = 0.84$).

Optimism

The effect of age was not significant for optimism, $F(1, 996) = 1.78$, $p = 0.182$. The results showed a significant main effect of gender role self-concept, $F(3, 996) = 29.47$, $p < 0.001$, $\eta^2p = 0.082$. A Bonferroni *post hoc* test showed that androgynous adolescents ($M = 3.65$, $SD = 0.81$) reported significantly higher scores on optimism than masculine ($M = 3.33$, $SD = 0.85$), feminine ($M = 3.35$, $SD = 0.79$), and undifferentiated ($M = 3.02$, $SD = 0.85$) adolescents, $p < 0.05$. Undifferentiated adolescents reported lower scores than all other groups, while there was no significant difference between masculine and feminine adolescents. The main effect of sex was not significant, $F(1, 996) = 1.32$, $p = 0.251$.

Connectedness

The results showed no significant effect of age for connectedness, $F(1, 996) < 1$, $p = 0.631$. The main effect of gender role self-concept was significant, $F(3, 996) = 32.73$, $p < 0.001$, $\eta^2p = 0.090$. A Bonferroni *post hoc* test revealed that androgynous adolescents ($M = 4.24$, $SD = 0.70$) reported significantly higher scores on connectedness than masculine ($M = 4.02$, $SD = 0.71$), feminine ($M = 3.96$, $SD = 0.83$), and undifferentiated ($M = 3.64$, $SD = 0.85$) adolescents, $p < 0.05$. Undifferentiated adolescents reported lower scores than all other groups, while there was no significant difference between masculine and feminine adolescents on connectedness. The main effect of sex was also significant, $F(1, 996) = 30.60$, $p < 001$, $\eta^2p = 0.030$. Girls ($M = 4.11$, $SD = 0.78$) scored significantly higher on connectedness than boys ($M = 3.82$, $SD = 0.80$).

Happiness

The effect of the covariate age was significant for happiness, $F(1, 996) = 31.54$, $p < 0.001$, $\eta^2p = 0.031$. There was a negative relationship between age and happiness, $r(998) = -0.167$, $p < 0.001$. The main effect of gender role self-concept was also significant

after controlling for age, $F(3, 996) = 31.74$, $p < 0.001$, $\eta^2 p = 0.088$. A Bonferroni *post hoc* test revealed that androgynous adolescents ($M = 3.81$, $SD = 0.86$) scored significantly higher on happiness than masculine ($M = 3.45$, $SD = 0.88$), feminine ($M = 3.39$, $SD = 0.87$), and undifferentiated ($M = 3.15$, $SD = 0.87$) adolescents, $p < 0.05$. Undifferentiated adolescents reported lower scores than all other groups, while there was no significant difference between masculine and feminine adolescents. The main effect of sex was significant after controlling for age, $F(1, 996) = 7.37$, $p < 0.01$, $\eta^2 p = 0.007$, with girls ($M = 3.53$, $SD = 0.88$) scoring higher on the happiness scale than boys ($M = 3.37$, $SD = 0.93$).

DISCUSSION

The goal of this study was to investigate relations between adolescent boys' and girls' gender role self-concepts and school-related well-being, taking into account both hedonic and eudemonic aspects of well-being. In general, our results support the androgyny model of well-being: androgynous boys and girls exhibited the highest levels of overall school-related well-being as well as the highest scores in all individual EPOCH-S dimensions: Engagement, Perseverance, Optimism, Connectedness, and Happiness. This finding is in line with Bem's theorizing that individuals who score high in both masculinity and femininity display better adjustment and greater psychological health (Bem, 1981, 1993), as well as other studies confirming this notion in adolescent samples (Boldizar, 1991; Pauletti et al., 2017). School is the first social space after the home in which individuals experience obligations, engagement, commitment and relationships. Thus, having a broader set of attributes and behavioral options that allow for flexible behavior and successful coping with different demands is more important in school than anywhere else during adolescence. This might be even more important in secondary school when classes become more challenging, peer relationships grow more complex, and educational and professional goals are developed and shaped (Brown and Larson, 2009; Verhoeven et al., 2019). Our results also confirmed our assumption that undifferentiated adolescents exhibit the lowest scores in overall school-related well-being as well as all individual dimensions. Similar to androgynous persons, undifferentiated individuals are not gender-typed (Bem, 1977), but unlike androgynous individuals, they lack the enriched behavioral repertoire of androgynous persons and at the same time do not possess the positive characteristics typical of either a masculine or feminine self-concept. For that reason, the lowest levels of well-being were found in this group in previous studies focusing on general well-being (see Markstrom-Adams, 1989 for review) and were expected in the context of school-related well-being as well.

However, contrary to our expectations, androgynous girls and boys scored higher on all individual EPOCH-S dimensions compared to masculine and feminine girls and boys. This finding is not surprising for the two hedonic dimensions of school-related well-being—optimism and happiness—which refer to positive affect in the school context and, as such, are not related to either masculinity or femininity. However, the eudemonic dimensions—engagement, perseverance and connectedness—have more gendered connotations and encompass instrumental and expressive goals and behaviors. With engagement and perseverance being clearly achievement-oriented, and connectedness reflecting social-related goals, we expected masculine adolescents to achieve the same results as androgynous adolescents in engagement and perseverance, and feminine adolescents to be similar to androgynous adolescents on connectedness. The non-significant differences between adolescents with masculine and feminine gender role self-concepts in hedonic, but especially eudemonic dimensions, as well as in overall school-related well-being, speak in favor of the equal importance of masculinity and femininity for school-related well-being. Despite the clear dominance of the masculinity model over the femininity model of well-being in the literature (Wichstrøm, 1999; Priess et al., 2009; Abele et al., 2016), femininity seems to be as relevant as masculinity for adolescents' functioning in school. This is not surprising giving the vital role of peer relationships in adolescence, as social interactions with classmates have been found to contribute to adolescents' well-being (e.g., Sandstrom and Dunn, 2014). Moreover, feminine students are more liked by teachers and obtain better grades (Heyder and Kessels, 2013), have stronger school-related self-esteem and exhibit stronger feelings of belonging at school (Skinner et al., 2019). Thus, our results confirm previous studies indicating the stronger importance of femininity in this period and in the school context compared to later in life (Heyder and Kessels, 2013; Chen et al., 2018).

The salience of femininity for school-related well-being can be explained with William James's theorizing that "self-centrality breeds self-enhancement," according to which people's judgement of their own self-worth is determined by the self-centrality of expressive or instrumental traits (Gebauer et al., 2013). Given the social nature of school, it can be argued that femininity occupies a more central position for adolescents' school-related well-being than for general well-being or well-being in adulthood. On the other hand, schooling continually emphasizes performance- and achievement-related goals in the form of competence-related feedback, performance-based evaluations and expectations of adolescents to be successful and ambitious about their futures, making instrumentality integral to school-related well-being as well. Chen et al. (2018) argued that expressivity is relatively more central to younger children's self-esteem due to their social goals, whereas instrumentality is relatively more central to older children's self-esteem due to the importance of performance-related goals in that period of life. The effects of age on school-related well-being in our study did not confirm this. The results showed a negative relationship between age and the "instrumental" dimension of the EPOCH-S measure, with younger adolescents scoring higher on engagement and perseverance, whereas age did not have an effect on connectedness, indicating equal endorsement of items related to connectedness among adolescents of all ages. Although femininity might be more central for self-esteem at a younger age, connectedness in school seems equally important across adolescence from the age of 12 to age 17. On the other hand, engagement and perseverance as defined in the EPOCH-S model seem to be more endorsed by younger adolescents. This is not surprising given that adolescents' behavioral and emotional involvement in academic activities declines as

they grow older (Archambault et al., 2009). Our results also show a negative relationship between age and overall school-related well-being and happiness. Younger adolescents exhibit higher overall well-being and happiness. This finding is consistent with previous studies showing a decrease in overall well-being and positive affect from early to middle adolescence (Goldbeck et al., 2007; González-Carrasco et al., 2017).

Although one might expect stereotypical gender differences in the eudemonic EPOCH-S dimensions, our results showed that girls score higher in engagement, perseverance, connectedness, happiness, and overall school-related well-being than boys. There were no significant differences between boys and girls in optimism. In studies with adults, sex differences were found in gendered dimensions of well-being, with men scoring higher than women in self-acceptance and autonomy and women scoring higher than men in positive relations with others (Matud et al., 2019). A study applying the original EPOCH model of adolescents' general well-being found small sex differences for optimism and connectedness only (Kern et al., 2016), with girls scoring higher on connectedness and boys on optimism. However, items assessing the EPOCH-S model are specific to the academic context, where girls exhibit higher engagement and tend to outperform boys (Duckworth and Seligman, 2006), which could explain girls' higher endorsement of EPOCH-S items. More importantly, gender role self-concept has been shown to be a more important determinant of well-being and psychological adjustment than biological sex (e.g., Bem, 1993; Priess et al., 2009; Chen et al., 2018). Sex differences in this context are therefore secondary to the effects of gender role self-concept. Moreover, the interaction between sex and gender role self-concept was not significant for overall school-related well-being or for any individual dimensions, supporting Bem's argument that androgyny does not offer more benefits to one gender than the other (Bem, 1993). In sum, our findings revealed that, beyond the effects of age and sex, androgynous adolescents experienced the highest levels of school-related well-being.

The results of this study have practical implications for school functioning. Developing curricular activities and a classroom environment that enhance both expressive and instrumental traits and behaviors in boys and girls may increase their school-related well-being. Interventions might focus on building performance- and achievement-related traits and goals as well as social-related traits and goals among all students, especially those at risk in terms of their well-being in school. These programs can not only contribute to school-related well-being among adolescents, but also decrease gender typing, which can result in reduced gender stereotypes in newer generations of adolescents.

Limitations and Future Directions

Although the current study provides valuable insights into the relationship between gender role self-concept and school-related well-being, several limitations must be considered. First, this study focuses only on gender role self-concept, that is, the self-perceived possession of specific gender-stereotyped attributes, and does not take into account other facets of gender role identity. Future studies might investigate the relationship between androgyny and school-related well-being by assessing gender identity as overall felt gender typicality (see Egan and Perry, 2001). Second, although the effect sizes of gender role self-concept comparisons for overall well-being were moderate, the effect sizes for single EPOCH-S dimensions were small, limiting the practical relevance of the identified coefficients. Third, the inventory used to assess gender role self-concept in this study comprised only positive masculine and feminine attributes. Although it has been noted that *positive* androgyny results in higher psychological well-being (Woodhill and Samuels, 2003), future studies should investigate the relationship between negative gender-stereotyped attributes and well-being in the school context, which might provide new insights. Moreover, although in line with original instrument (Krahé et al., 2007), alpha reliabilities for both femininity and masculinity scale are rather low, indicating poor internal consistency of items constructing gender role self-concept types. However, scholars argue that constellations of gendered attributes similar to those used in our study reliably predict health and well-being (see, e.g., Abele, 2014; Yarnell et al., 2019). Future study thus should test similar design with another gender role identity measure for adolescent population. Fourth, this study does not address other factors impacting both gender socialization and school-related well-being. Further studies are needed to investigate the role of peers and teachers for the relation between gender role self-concept and school-related well-being. Finally, this study was conducted in a Western country and the results should not be generalized to other cultures. Androgyny seems to be beneficial for school-related well-being in individualist societies where traditional gender roles might be more liberal. In societies with more rigid gender norms or different cultural contexts, a different interplay between gender role self-concept and school-related well-being might be observed.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/5umnp/?view_only=2ea6ea1ceb61477192ea0f0fdb59aff1.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Participation in the study was completely voluntary. Only those who gave active consent took part. Additionally, parental consent was obtained for participation in the study as well as data usage.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Abele, A. E. (2014). Pursuit of communal values in an agentic manner: a way to happiness? *Front. Psychol.* 5:1320. doi: 10.3389/fpsyg.2014.01320

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., and Duan, Y. (2016). Facets of the fundamental content dimensions: agency with competence and assertiveness—communion with warmth and morality. *Front. Psychol.* 7:1810. doi: 10.3389/fpsyg.2016.01810

Abele, A. E., and Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *J. Pers. Soc. Psychol.* 93, 751–763. doi: 10.1037/0022-3514.93.5.751

Antaramian, S. P., Huebner, E. S., Hills, K. J., and Valois, R. F. (2010). A dual-factor model of mental health: toward a more comprehensive understanding of youth functioning. *Am. J. Orthopsychiatry* 80, 462–472. doi: 10.1111/j.1939-0025.2010.01049.x

Archambault, I., Janosz, M., Morizot, J., and Pagani, L. (2009). Adolescent behavioral, affective, and cognitive engagement in school: relationship to dropout. *J. Sch. Health* 79, 408–415. doi: 10.1111/j.1746-1561.2009.00428.x

Bem, S. L. (1974). The measurement of psychological androgyny. *J. Consult. Clin. Psychol.* 42, 155–162. doi: 10.1037/h0036215

Bem, S. L. (1977). On the utility of alternative procedures for assessing psychological androgyny. *J. Consult. Clin. Psychol.* 45, 196–205. doi: 10.1037/0022-006X.45.2.196

Bem, S. L. (1981). Gender schema theory: a cognitive account of sex typing. *Psychol. Rev.* 88, 354–364. doi: 10.1037/0033-295X.88.4.354

Bem, S. L. (1993). *The Lenses of Gender—Transforming the Debate on Sexual Inequality*. New Haven: Yale University Press.

Berger, C., Alcalay, L., Torretti, A., and Milicic, N. (2011). Socio-emotional well-being and academic achievement: evidence from a multilevel approach. *Psicologia: Reflexão e Crítica* 24, 344–351. doi: 10.1590/S0102-79722011000200016

Boldizar, J. P. (1991). Assessing sex typing and androgyny in children: the children's sex role inventory. *Dev. Psychol.* 27, 505–515. doi: 10.1037/0012-1649.27.3.505

Brown, B. B., and Larson, J. (2009). "Peer relationships in adolescence," in *Handbook of Adolescent Psychology*. eds. R. M. Lerner and L. Steinberg (New York: John Wiley and Sons, Inc).

Buchanan, K., and Bardi, A. (2015). The roles of values, behavior, and value-behavior fit in the relation of agency and communion to well-being: agency, communion, and well-being. *J. Pers.* 83, 320–333. doi: 10.1111/jopy.12106

Buerger, S., Holzer, J., Yanagida, T., Schober, B., and Spiel, C. (2022). The German Version of the EPOCH Measure of Adolescent Well-Being Adapted for the School Context (EPOCH-G-S) -A Validation Study. Manuscript submitted for publication.

Carver, C. S., Scheier, M. F., and Segerstrom, S. C. (2010). Optimism. *Clin. Psychol. Rev.* 30, 879–889. doi: 10.1016/j.cpr.2010.01.006

Carver, L. F., Vafaei, A., Guerra, R., Freire, A., and Phillips, S. P. (2013). Gender differences: examination of the 12-item bem sex role inventory (BSRI-12) in an older Brazilian population. *PLoS One* 8:e76356. doi: 10.1371/journal.pone.0076356

Carver, P. R., Yunger, J. L., and Perry, D. G. (2003). Gender identity and adjustment in middle childhood. *Sex Roles* 49, 95–109. doi: 10.1023/A:1024423012063

Chen, F., Zhu, S., and Bi, C. (2018). The development of self-esteem and the role of agency and communion: a longitudinal study among Chinese. *J. Child Fam. Stud.* 27, 816–824. doi: 10.1007/s10826-017-0942-y

Choudhury, S., Blakemore, S.-J., and Charman, T. (2006). Social cognitive development during adolescence. *Soc. Cogn. Affect. Neurosci.* 1, 165–174. doi: 10.1093/scan/nsl024

Corby, B. C., Hodges, E. V. E., and Perry, D. G. (2007). Gender identity and adjustment in black, Hispanic, and white preadolescents. *Dev. Psychol.* 43, 261–266. doi: 10.1037/0012-1649.43.1.261

DiDonato, M. D., and Berenbaum, S. A. (2011). The benefits and drawbacks of gender typing: how different dimensions are related to psychological adjustment. *Arch. Sex. Behav.* 40, 457–463. doi: 10.1007/s10508-010-9620-5

Duckworth, A. L., and Seligman, M. E. P. (2006). Self-discipline gives girls the edge: gender in self-discipline, grades, and achievement test scores. *J. Educ. Psychol.* 98, 198–208. doi: 10.1037/0022-0663.98.1.198

Eccles, J. S. (2004). "Schools, academic motivation, and stage-environment fit," in *Handbook of Adolescent Psychology*. eds. R. M. Lerner and L. Steinberg (New York: John Wiley and Sons Inc), 125–153.

Eccles, J. S., and Roeser, R. W. (2011). Schools as developmental contexts during adolescence: schools as developmental contexts. *J. Res. Adolesc.* 21, 225–241. doi: 10.1111/j.1532-7795.2010.00725.x

Egan, S. K., and Perry, D. G. (2001). Gender identity: a multidimensional analysis with implications for psychosocial adjustment. *Dev. Psychol.* 37, 451–463. doi: 10.1037/0012-1649.37.4.451

Elmore, G. M., and Huebner, E. S. (2010). Adolescents' satisfaction with school experiences: relationships with demographics, attachment relationships, and school engagement behavior. *Psychol. Sch.* 47, 525–537. doi: 10.1002/pits.20488

Gebauer, J. E., Wagner, J., Sedikides, C., and Neberich, W. (2013). Agency-communion and self-esteem relations are moderated by culture, religiosity, age, and sex: evidence for the "self-centrality breeds self-enhancement," principle: agency-communion and self-esteem. *J. Pers.* 81, 261–275. doi: 10.1111/j.1467-6494.2012.00807.x

Gilman, R., and Huebner, E. S. (2006). Characteristics of adolescents who report very high life satisfaction. *J. Youth Adolesc.* 35, 293–301. doi: 10.1007/s10964-006-9036-7

Goldbeck, L., Schmitz, T. G., Besier, T., Herschbach, P., and Henrich, G. (2007). Life satisfaction decreases during adolescence. *Qual. Life Res.* 16, 969–979. doi: 10.1007/s11136-007-9205-5

González-Carrasco, M., Casas, F., Malo, S., Viñas, F., and Dinisman, T. (2017). Changes with age in subjective well-being through the adolescent years: differences by gender. *J. Happiness Stud.* 18, 63–88. doi: 10.1007/s10902-016-9717-1

Helgeson, V. S., and Palladino, D. K. (2012). Implications of psychosocial factors for diabetes outcomes among children with type 1 diabetes: a review: psychosocial factors in diabetes. *Soc. Personal. Psychol. Compass* 6, 228–242. doi: 10.1111/j.1751-9004.2011.00421.x

Heyder, A., and Kessels, U. (2013). Is school feminine? Implicit gender stereotyping of school as a predictor of academic achievement. *Sex Roles* 69, 605–617. doi: 10.1007/s11199-013-0309-9

Hill, J. P., and Lynch, M. E. (1983). "The intensification of gender-related role expectations during early adolescence," in *Girls at Puberty*. eds. J. Brooks-Gunn and A. C. Petersen (Berlin: Springer), 201–228.

Holzer, J., Bürger, S., Samek-Krenkel, S., Spiel, C., and Schober, B. (2021). Conceptualization of students' school-related wellbeing: students' and teachers' perspectives. *Educ. Res.* 63, 474–496. doi: 10.1080/00131881.2021.1987152

Howell, A. J. (2009). Flourishing: achievement-related correlates of students' well-being. *J. Posit. Psychol.* 4, 1–13. doi: 10.1080/17439760802043459

Jourdan, D., Samdal, O., Diagne, F., and Carvalho, G. S. (2008). The future of health promotion in schools goes through the strengthening of teacher training at a global level. *Promot. Educ.* 15, 36–38. doi: 10.1177/1025382308095657

Kahneman, D. (1999). "Objective happiness," in *Well-Being: The Foundations of Hedonic Psychology*. eds. D. Kahneman, E. Diener and N. Schwarz (New York: Russell Sage Foundation), 3–25.

Kern, M. L., Benson, L., Steinberg, E. A., and Steinberg, L. (2016). The EPOCH measure of adolescent well-being. *Psychol. Assess.* 28, 586–597. doi: 10.1037/pas0000201

Kern, M. L., Waters, L. E., Adler, A., and White, M. A. (2015). A multidimensional approach to measuring well-being in students: application of the PERMA framework. *J. Posit. Psychol.* 10, 262–271. doi: 10.1080/17439760.2014.936962

Klaczynski, P. A., Felmban, W. S., and Kole, J. (2020). Gender intensification and gender generalization biases in pre-adolescents, adolescents, and emerging adults. *Br. J. Dev. Psychol.* 38, 415–433. doi: 10.1111/bjdp.12326

Krahé, B., Berger, A., and Möller, I. (2007). Entwicklung und Validierung eines Inventars zur Erfassung des Geschlechtsrollen-Selbstkonzepts im Jugendalter. *Zeitschrift Fur Sozialpsychologie - Z SOZPSYCHOL* 38, 195–208. doi: 10.1024/0044-3514.38.3.195

Laski, L. (2015). Realizing the health and wellbeing of adolescents. *BMJ* 351:h4119. doi: 10.1136/bmj.h4119

Lewis, A. D., Huebner, E. S., Malone, P. S., and Valois, R. F. (2011). Life satisfaction and student engagement in adolescents. *J. Youth Adolesc.* 40, 249–262. doi: 10.1007/s10964-010-9517-6

Liu, W., Mei, J., Tian, L., and Huebner, E. S. (2016). Age and gender differences in the relation between school-related social support and subjective well-being in school among students. *Soc. Indic. Res.* 125, 1065–1083. doi: 10.1007/s11205-015-0873-1

Long, R. F., and Huebner, E. S. (2014). Differential validity of global and domain-specific measures of life satisfaction in the context of schooling. *Child Indic. Res.* 7, 671–694. doi: 10.1007/s12187-013-9231-5

Long, R. F., Huebner, E. S., Wedell, D. H., and Hills, K. J. (2012). Measuring school-related subjective well-being in adolescents. *Am. J. Orthopsychiatry* 82, 50–60. doi: 10.1111/j.1939-0025.2011.01130.x

Markstrom-Adams, C. (1989). Androgyny and its relation to adolescent psychosocial well-being: a review of the literature. *Sex Roles J. Res.* 21, 325–340. doi: 10.1007/BF00289595

Martínez-Marín, M. D., and Martínez, C. (2019). Subjective well-being and gender-typed attributes in adolescents: The relevance of emotional intelligence. *Aust. J. Psychol.* 71, 296–304. doi: 10.1111/ajpy.12247

Matud, M. P., López-Curbelo, M., and Fortes, D. (2019). Gender and psychological well-being. *Int. J. Environ. Res. Public Health* 16, 3531. doi: 10.3390/ijerph16193531

McKnight, C. G., Huebner, E. S., and Suldo, S. (2002). Relationships among stressful life events, temperament, problem behavior, and global life satisfaction in adolescents. *Psychol. Sch.* 39, 677–687. doi: 10.1002/pits.10062

Menon, M., Schellhorn, K., and Lowe, C. A. (2013). Interactive influences of gender identity and gender typing on early adolescents' well-being. *Psychol. Stud.* 58, 1–9. doi: 10.1007/s12646-012-0159-5

Park, N. (2004). The role of subjective well-being in positive youth development. *Ann. Am. Acad. Pol. Soc. Sci.* 591, 25–39. doi: 10.1177/0002716203260078

Pauletti, R. E., Menon, M., Cooper, P. J., Aults, C. D., and Perry, D. G. (2017). Psychological androgyny and children's mental health: a new look with new measures. *Sex Roles* 76, 705–718. doi: 10.1007/s11199-016-0627-9

Pedhazur, E. J., and Tetenbaum, T. J. (1979). Bem sex role inventory: a theoretical and methodological critique. *J. Pers. Soc. Psychol.* 37, 996–1016. doi: 10.1037/0022-3514.37.6.996

Priess, H. A., Lindberg, S. M., and Hyde, J. S. (2009). Adolescent gender-role identity and mental health: gender intensification revisited. *Child Dev.* 80, 1531–1544. doi: 10.1111/j.1467-8624.2009.01349.x

Pyhältö, K., Soini, T., and Pietarinen, J. (2010). Pupils' pedagogical well-being in comprehensive school—significant positive and negative school experiences of Finnish ninth graders. *Eur. J. Psychol. Educ.* 25, 207–221. doi: 10.1007/s10212-010-0013-x

Ryan, R. M., and Deci, E. L. (2001). On happiness and human potentials: a review of research on hedonic and Eudaimonic well-being. *Annu. Rev. Psychol.* 52, 141–166. doi: 10.1146/annurev.psych.52.1.141

Sandstrom, G. M., and Dunn, E. W. (2014). Social interactions and well-being: the surprising power of weak ties. *Personal. Soc. Psychol. Bull.* 40, 910–922. doi: 10.1177/0146167214529799

Seligman, M. E. P. (2011). *Flourish: A Visionary New Understanding of Happiness and Well-Being*. New York: Free Press.

Seligman, M. E. P., and Csikszentmihalyi, M. (2000). Positive psychology: an introduction. *Am. Psychol.* 55, 5–14. doi: 10.1037/0003-066X.55.1.5

Skinner, O. D., McHale, S. M. D., and Telfer, N. A. (2019). Gender-typed personality qualities and African American youth's school functioning. *J. Youth Adolesc.* 48, 680–691. doi: 10.1007/s10964-018-0919-1

Spence, J. T. (1991). Do the Bsri and Paq measure the same or different concepts? *Psychol. Women Q.* 15, 141–165. doi: 10.1111/j.1471-6402.1991.tb00483.x

Spence, J. T., and Helmreich, R. L. (1979). Comparison of masculine and feminine personality attributes and sex-role attitudes across age groups. *Dev. Psychol.* 15, 583–584. doi: 10.1037/h0078091

Spence, J. T., Helmreich, R., and Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *J. Pers. Soc. Psychol.* 32, 29–39. doi: 10.1037/h0076857

Steinmayr, R., Heyder, A., Naumburg, C., Michels, J., and Wirthwein, L. (2018). School-related and individual predictors of subjective well-being and academic achievement. *Front. Psychol.* 9:2631. doi: 10.3389/fpsyg.2018.02631

Tian, L., Chen, H., and Huebner, E. S. (2014). The longitudinal relationships between basic psychological needs satisfaction at school and school-related subjective well-being in adolescents. *Soc. Indic. Res.* 119, 353–372. doi: 10.1007/s11205-013-0495-4

Verhoeven, M., Poorthuis, A. M. G., and Volman, M. (2019). The role of school in adolescents' identity development. A literature review. *Educ. Psychol. Rev.* 31, 35–63. doi: 10.1007/s10648-018-9457-3

Wang, Y. N. (2016). Balanced authenticity predicts optimal well-being: theoretical conceptualization and empirical development of the authenticity in relationships scale. *Personal. Individ. Differ.* 94, 316–323. doi: 10.1016/j.paid.2016.02.001

Whitley, B. E. (1985). Sex-role orientation and psychological well-being: two meta-analyses. *Sex Roles* 12, 207–225. doi: 10.1007/BF00288048

Wichstrøm, L. (1999). The emergence of gender difference in depressed mood during adolescence: The role of intensified gender socialization. *Dev. Psychol.* 35, 232–245. doi: 10.1037/0012-1649.35.1.232

Wolfram, H.-J., Mohr, G., and Borchert, J. (2009). Gender role self-concept, gender-role conflict, and well-being in male primary school teachers. *Sex Roles* 60, 114–127. doi: 10.1007/s11199-008-9493-4

Woodhill, B. M., and Samuels, C. A. (2003). Positive and negative androgyny and their relationship with psychological health and well-being. *Sex Roles J. Res.* 48, 555–565. doi: 10.1023/A:1023531530272

Yang, Q., Tian, L., Huebner, E. S., and Zhu, X. (2018). Relations among academic achievement, self-esteem, and subjective well-being in school among elementary school students: a longitudinal mediation model. *School Psychol. Q.* 34, 328–340. doi: 10.1037/spq0000292

Yarnell, L. M., Neff, K. D., Davidson, O. A., and Mullarkey, M. (2019). Gender differences in self-compassion: examining the role of gender role orientation. *Mindfulness* 10, 1136–1152. doi: 10.1007/s12671-018-1066-1

Yunger, J. L., Carver, P. R., and Perry, D. G. (2004). Does gender identity influence children's psychological well-being? *Dev. Psychol.* 40, 572–582. doi: 10.1037/0012-1649.40.4.572

Žukauskienė, R. (2014). "Adolescence and well-being," in *Handbook of Child Well-Being*. eds. A. Ben-Arieh, F. Casas, I. Frønes and J. E. Korbin (Berlin: Springer), 1713–1738.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 27 May 2022
doi: 10.3389/fpsyg.2022.821227

Influence of Flexible Classroom Seating on the Wellbeing and Mental Health of Upper Elementary School Students: A Gender Analysis

Jonathan Bluteau[*†], *Solène Aubenas*[†] *and France Dufour*

Département d'Éducation et Formation Spécialisées, Université du Québec à Montréal, Montréal, QC, Canada

While traditional seating (also known as *fixed seating* or *fixed classroom*) remains the preferred classroom seating arrangement for teachers, a new type of seating arrangement is becoming more common in schools: the flexible classroom (also known as *flexible seating*). The purpose of this type of arrangement is to meet the needs of students by providing a wide variety of furniture and workspaces, to put students at the center of learning, and to allow them to make choices based on their preferences and the objectives of the task at hand. This study aimed to examine the influence of flexible seating on the wellbeing and mental health of elementary school students. This article presents the results of exploratory research conducted in Quebec among Grade 5 and 6 students comparing the wellbeing and mental health of students in fixed and flexible classrooms. The study was conducted with 107 students in three Grade 5 and 6 flexible classrooms ($n = 51$) and three Grade 5 and 6 fixed classrooms ($n = 56$). It is based on a quasi-experimental, quantitative design with post-test only and a control group. The groups were matched based on natural conditions (i.e., from a convenience sample). Furthermore, the study included a gender-differentiated analysis for each group. The results showed that flexible classroom seating had a positive influence on the girls' wellbeing and mental health. In contrast, for the boys, fixed classroom seating was most conducive to their wellbeing and mental health. However, our study has some limitations that are discussed in the article.

Keywords: physical environment, classroom layout, flexible seating, wellbeing, mental health, upper elementary school

OPEN ACCESS

Edited by:
Caterina Fiorilli,
Libera Università Maria SS. Assunta University, Italy

Reviewed by:
Eleonora Topino,
Libera Università Maria SS. Assunta University, Italy
Lidia Scifo,
Libera Università Maria SS. Assunta, Italy

***Correspondence:**
Jonathan Bluteau
bluteau.jonathan@uqam.ca

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *24 November 2021*
Accepted: *21 April 2022*
Published: *27 May 2022*

Citation:
Bluteau J, Aubenas S and Dufour F (2022) Influence of Flexible Classroom Seating on the Wellbeing and Mental Health of Upper Elementary School Students: A Gender Analysis. Front. Psychol. 13:821227. doi: 10.3389/fpsyg.2022.821227

INTRODUCTION

Over the past few years, there has been a growing awareness in the education community about the importance of the school physical environment (Gouvernement du Québec. Ministère de l'Éducation et de l'Enseignement Supérieur, 2020). Indeed, by the possibility that students have to interact with the physical school environment through movement, exploration, and social interaction, it would strengthen the physical, cognitive, emotional, and social development of students based on the principle that a well thought-out school physical environment promotes the global development of students, promotes academic wellbeing and inclusion (Aziz et al., 2017). In fact, the fields of architecture and educational psychology have looked at the dimensions of the physical environment

that can impact on the global development of students. According to an interdisciplinary perspective, the school space is thought of as a living environment where the student and his environment interact in a transactional way and mutually define each other (Jodelet, 2015). In this perspective, the theoretical model proposed by Pianta et al. (2008), on transactional and developmental theories, presents an interactional model between students, the teacher, and the school environment. The school environment is considered in all aspects of the daily experience and interactions that the student has with his teacher and his peers. As a result, the educational environment influences the resulting social interactions and the cognitive and socio-affective development of students (Bronfenbrenner and Morris, 1998; Broto, 2013; Huynh et al., 2013).

Numerous international publications (Organisation de coopération et de développement économiques, 2001, 2011) on the 21st-century school were pivotal in considering the role of the physical environment in students' school experiences. As a result, many education systems, such as those in Quebec, France, Germany, Denmark, and Finland, have begun to ask whether the school environment in which students develop can contribute to their sense of wellbeing and, ultimately, to their success. At the same time, teachers are increasingly interested in the question of how to structure their classrooms to meet teaching requirements and support learning. The term "classroom physical environment" refers to all the furniture and its spatial arrangement in the classroom (Abbasi, 2013). We note that the majority of classroom arrangements, particularly in Quebec, remain fixed classrooms. However, in recent years, a new type of classroom seating arrangement has developed: the flexible classroom seating (Laquerre, 2018; Vallée, 2019).

The fixed classroom, also known as the traditional classroom, is the most commonly observed seating arrangement in schools and is also associated with teacher-centered practice. In this type of classroom, there are as many desks as there are students, and the teacher is usually responsible for assigning a desk to each student. Desk arrangement can vary—rows, U-shaped, or clusters. There are several explanations for the choice of desk arrangement. Desks arranged in rows especially encourage individual work (Wannarka and Ruhl, 2008), while U-shaped or cluster arrangements encourage social interaction and cooperation (Wannarka and Ruhl, 2008; Farmer et al., 2011; Gest and Rodkin, 2011). However, desk arrangement is rarely changed during the year, and classrooms are not routinely rearranged for a particular teaching activity.

The second type of arrangement, called the flexible classroom seating, is currently gaining traction with teachers (Laquerre, 2018; Vallée, 2019). In Quebec, it is estimated that there are more than 1,500 flexible classrooms in place throughout its School Services Centres (CSSs; Bluteau et al., 2019). Adoption of the flexible classroom has spread as a result of social networks (Havig, 2017). In this type of classroom, some or all of the desks have been replaced by a wide range of so-called flexible furniture that offers a variety work surfaces, seating sizes and heights, body positions (Dornfeld, 2016; Havig, 2017; Limpert, 2017; Del'Homme, 2018; Laquerre, 2018; Legout, 2018; Tiennot, 2019; Vallée, 2019). In this way, students do not have assigned seating (Legout, 2018). They can move about in the classroom and choose the seat that best suits them for the task at hand (Dornfeld, 2016; Havig, 2017; Limpert, 2017; Del'Homme, 2018; Laquerre, 2018; Legout, 2018; Tiennot, 2019; Vallée, 2019). It allows students to explore, move about, experiment, manipulate, and make the space and furniture their own, with the goal of encouraging original and creative ways of experiencing the classroom (Abbasi, 2013; Mazalto and Paltrinieri, 2013; Keymeulen et al., 2020). Flexible furniture is also designed so that classrooms can be modified easily. Teachers can therefore rearrange their classroom to suit the teaching activity and the type of behavior expected (Wannarka and Ruhl, 2008; Havig, 2017; Carignan, 2018; Erz, 2018; Keymeulen et al., 2020), including group work, pair work, or individual work. From this perspective, this type of classroom arrangement allows for implementing teaching practices that can be described as "flexible," that is, student-centered, differentiated, and collaborative (Barrett et al., 2015, 2017; Delzer, 2015; Dornfeld, 2016; Havig, 2017; Erz, 2018; Keymeulen et al., 2020). Flexible classrooms address the principles of the "Current pedagogical discourse […] focused on learning, on putting the student at the center of the discussion, on helping them to be adaptive, creative, cooperative, responsive, and self-reliant."[1] (Blyth, 2013, p. 53). Thus, the flexible classroom seating is more associated with student-centered teaching practice.

Previous research has found that the functioning of the flexible classroom seating contributes to the development of certain personal skills, such as self-reliance, self-regulation, and problem-solving (Doyon, 2018; Erz, 2018; Laquerre, 2018; Legout, 2018). The flexible seating can therefore help to empower students and make them actors in their own learning (Legout, 2018). Furthermore, the functioning of the flexible classroom seating can have positive effects particularly on attention, motivation, engagement, and the adoption of task-appropriate behavior (Delzer, 2015; Dornfeld, 2016; Boudreault, 2017; Comaianni, 2017; Limpert, 2017; Allen, 2018; Erz, 2018; Laquerre, 2018; Legout, 2018; Schrage, 2018; Tiennot, 2019). By encouraging movement, choice, and interaction, and increasing students' sense of control, this type of classroom arrangement addresses students' physical, social, and cognitive needs (Comaianni, 2017; Havig, 2017; Limpert, 2017; Erz, 2018; Legout, 2018; Schoolcraft, 2018; Schrage, 2018; Sorrell, 2019; Vallée, 2019). However, there are limitations to flexible seating addressed in the literature that should be mentioned. On the one hand, this type of classroom arrangement may be challenging for students who need guidance and routine (Legout, 2018; Vallée, 2019). On the other hand, some students may be challenged by the lack of personal space. As Legout (2018) found, shared furniture and space, and no longer having an assigned desk, do not work for all students. Nevertheless, despite the growing interest by teachers in this

Abbreviations: CIEREH, Institutional Human Research Committee (Comité institutionnel d'éthique de la recherche avec des êtres humains); CSS, School Services Centre (Centre de services scolaire); CSE, Higher Education Council (Conseil Supérieur de l'Éducation); SD, Standard Deviation; IMSE, Socio-economic Background Index (Indices du milieu socio-économique); M, Mean.

[1]"discours pédagogique actuel […] axé sur l'apprentissage, sur le fait de mettre l'élève au centre du débat, de l'aider à s'adapter, à être créatif, coopératif, réactif et autonome." [Our translation].

type of arrangement, the flexible classroom remains poorly documented in the literature (Havig, 2017; Laquerre, 2018; Vallée, 2019), and its influence on student learning and mental health is still poorly understood. As a result, there is a sizeable gap between research that is currently available and the enthusiasm that this type of arrangement has generated among teachers (Havig, 2017; Laquerre, 2018; Vallée, 2019).

Furthermore, interest in school wellness and mental health came late to the field of studies in education (Piché et al., 2017) and has become prominent in many educational system reforms (Bacro et al., 2017). Indeed, the redefinition of success to include various aspects of students' holistic development has made wellbeing a fundamental concept of the 21st-century school (Guimard et al., 2015; Ferrière et al., 2016). Numerous studies have revealed that lived school experiences are associated with development, identity construction, academic success, and wellbeing (Konu and Rimpelä, 2002; Wigfield et al., 2006; Eccles and Roeser, 2011; Rousseau, 2012; Bacro et al., 2017). In this sense, wellbeing at school may depend on many factors rooted in students' school experience (Guimard et al., 2015, as cited in Fouquet-Chauprade, 2013; Ferrière et al., 2016). Wellbeing is a multidimensional, multifactorial, and systemic concept (Conseil Supérieur de l'Éducation, 2020). It has been characterized according to objective, subjective, environmental, and contextual factors (Espinosa and Rousseau, 2018). The scientific literature takes two divergent paths to define wellbeing. On the one hand, the hedonic conception associates wellbeing with pleasure, satisfaction, and subjective happiness (Laguardia and Ryan, 2000; Doré and Caron, 2017; Conseil Supérieur de l'Éducation, 2020). Thus, a positive sense of wellbeing "consists of experiencing many positive affects, few unpleasant ones" (Laguardia and Ryan, 2000, p. 282), but also "feeling a high overall satisfaction with one's life"[2] (Florin and Guimard, 2017, p. 20). On the other hand, the eudemonic conception of wellbeing refers to personal fulfillment and self-actualization (Laguardia and Ryan, 2000; Conseil Supérieur de l'Éducation, 2020). Commonly referred to as psychological wellbeing, this conception is more recent (Antoine et al., 2007). Here, wellbeing consists solely in living in accordance with one's own nature and values (Laguardia and Ryan, 2000; Conseil Supérieur de l'Éducation, 2020). For a long time, these two approaches have represented divergent directions for research. However, at present, a combination of the two conceptions would seem necessary to encompass wellbeing in its entirety: "Well-being should be understood as a state of subjective pleasure and satisfaction with life, but also of self-actualization"[3] (Conseil Supérieur de l'Éducation, 2020, p. 20). This more encompassing definition comes close to the current World Health Organization (WHO) definition of good mental health. Indeed, the rise of positive psychology has led to a more encompassing definition of mental health (Ferrière et al., 2016; Doré and Caron, 2017; Shankland et al., 2017). This new branch in psychology is defined as, "the scientific study of positive experiences, wellbeing, and optimal functioning of the individual"[4] (Antoine et al., 2007, p. 170). The definition of positive mental health, also known as optimal mental health, takes into account the wellbeing and good psychological and social functioning of the individual (Doré and Caron, 2017; Shankland et al., 2017; Conseil Supérieur de l'Éducation, 2020). Indeed, mental health includes all dimensions of a student's overall development (Welsh et al., 2015) and can be defined by low stress, a sense of psychological wellbeing, and ultimately, good coping and behavioral functioning. Thus, mental health and wellbeing are closely linked (Conseil Supérieur de l'Éducation, 2020). Protective factors that positively influence mental health and wellbeing in school and decrease exposure to stressors include the quality of the physical environment, classroom interactions (Amoly et al., 2014), and social support provided by the teacher (Kruger et al., 2007; Heaney and Israel, 2008). However, the wellbeing, and ultimately, the mental health of students can be influenced by the quality of the environments they occupy, which can be explained by "the degree to which psychological and/or physiological needs are met in each environment, the physical perception of the environment, and the atmosphere of the environment" (Joing et al., 2018, p. 19). Furthermore, few studies to our knowledge have examined the relationship between classroom seating arrangement and student wellbeing and mental health. To date, the few studies that have examined the effects of seating arrangement show that certain aspects of the school physical environment (natural light, space, air quality, and so on) are linked to student concentration in the classroom and academic achievement (Cheryan et al., 2014; Barrett et al., 2017). However, in studies related to school architecture, little attention has been given to student mental health and wellbeing in the classroom. Nevertheless, seating arrangement and furniture play a role in teaching situations and overall development, and may ultimately influence student wellbeing and mental health (Mazalto, 2017; Doyon, 2018; Erz, 2018; Joing et al., 2018; Laquerre, 2018; Legout, 2018). As such, examining the influence of seating arrangement on mental health and wellbeing by investigating two types of classroom arrangements (fixed classroom seating and flexible classroom seating) may fill a gap in the existing literature on the topic, which this study has sought to do.

With this in mind, the question that guided this study was "Does classroom seating influence the academic wellbeing and mental health of elementary school students?" To answer this question, the study had two specific objectives: (1) compare the wellbeing and mental health of Grade 5 and 6 students in the context of differentiated classroom arrangements (fixed and flexible classrooms); and (2) compare, by gender, the wellbeing and mental health of Grade 5 and 6 students in the context of differentiated classroom arrangements (fixed and flexible classrooms).

Our starting hypothesis was to observe differences between the groups on the variables studied and an increase of wellbeing and better mental health in the flexible group, with no difference in terms of gender.

[2]"ressentir une grande satisfaction générale à l'égard de sa vie." [Our translation].
[3]"le bien-être devrait être compris comme un état de plaisir subjectif et de satisfaction à l'égard de la vie, mais aussi de réalisation de soi." [Our translation].
[4]"l'étude scientifique des expériences positives, du bien-être et du fonctionnement optimal de l'individu." [Our translation].

MATERIALS AND METHODS

Research Design

This study is part of the Social Sciences and Humanities Research Council-funded research project "Influence of Classroom Seating Arrangement and Quality of Teacher-Student Interactions on Stress Coping and School Mental Health of Elementary School Students" (Bluteau et al., 2019). The study was approved by the Institutional Human Research Ethics Committee (CIEREH) of the Université du Québec à Montréal (UQAM) in October 2019. The study bears the following ethics certificate number: 3761_e_2019. The CIEREH agreement was then sent to the head office of the School Services Centre (CSS) for verification and approval.

The study is based on a quasi-experimental, quantitative design with post-test only and a control group. The groups were matched based on natural conditions (i.e., from a convenience sample). Participants were Grade 5 and 6 students in three fixed and flexible classrooms. Wellbeing and mental health were studied in both groups, and for boys and girls in each group.

Participants

Teachers

Although teachers were not the focus of the study, teachers are in charge of the classroom and subject to its design. Thus, for research validity, it was important to select teachers in such a way as to control for teacher effect. Teacher effect can be perceived through the teacher's attitude, experience, and sense of efficacy, among others. For this reason, a questionnaire was given to each teacher who wanted to participate. The purpose of the questionnaire, whose variables will be described in Section "Measuring Instruments and Data Collection Procedures," was to match the three flexible classroom teachers to three fixed classroom teachers with similar teacher profiles in terms of sense of self-efficacy, job satisfaction, intention in the performance goal structure, years of teaching experience, and age. In the end, each flexible classroom teacher was matched to three fixed classroom teachers based on scale score equivalence. Consequently, six female teachers agreed to have students in their classrooms participate. The mean age of the female teachers in the fixed classrooms was 36 (SD = 6.36), and in the flexible classrooms (SD = 2.65) it was 34. As such, two fixed Grade 5 classrooms and one fixed Grade 6 classroom were matched to two flexible Grade 5 classrooms and one flexible Grade 6 classroom. In this way, each pair of teachers taught at the same level, had the same profile, and were in the same age range.

Students

The sample consisted of 107 students: 51 students in fixed classrooms (24 girls and 27 boys) and 56 students in flexible classrooms (26 girls and 30 boys). We note that the number of students in each group and the proportion of girls to boys in each group were relatively equal. The mean age of students in the fixed classrooms was 11.13 (SD = 0.54). Students in the flexible classrooms had a mean age of 11.23 (SD = 0.7; see **Table 1**).

TABLE 1 | Sample description.

	Total	Fixed group	Flexible group
Student data		51	56
Age; Mean (SD)	107	11.13 (0.5)	11.23 (0.7)
Girls	50	24	26
Boys	57	27	30

Procedures

To limit bias, the study ensured that the two groups of students were equivalent on multiple levels, namely, (1) the School Services Centre, (2) the teachers, (3) the classes, and (4) the students.

The first phase consisted of establishing a partnership with a Montreal South Shore School Services Centre (CSS). The six classes in the sample were drawn from the same CSS. Notably, the CSS had been undertaking numerous expansion and construction projects to redesign classrooms and schools. Furthermore, the CSS was implementing flexible classrooms in a controlled manner to document the testing of this type of classroom and to examine its impact on indicators of educational success.

In the second phase, initial contact was made with the teachers of the partner CSS. Approximately 50 CSS teachers were contacted by email. The objective was to recruit teachers interested in the project who were teaching Grade 5 and 6 students in flexible classrooms. They were asked if they were interested in the project and, if so, whether they taught in a flexible classroom environment with Grade 5 and 6 students. As a result, three flexible classroom teachers who met the criteria were selected. Subsequently, a CSS manager in charge of the study was able to provide us with contact information for eight Grade 5 and 6 teachers (one male and seven female teachers) who had opted for the fixed classroom arrangement. A matching questionnaire (see "Teacher Matching") was distributed to the three flexible classroom teachers and the eight fixed classroom teachers in order to pair the three flexible classroom teachers with three fixed classroom teachers based on a number of criteria. A total of seven fixed classroom teachers responded to the questionnaire, three of whom were matched to the three flexible classroom teachers.

Next, we examined the socio-economic background index for each school in our sample. The schools in our sample were public schools located in rural and semi-rural areas. The Disadvantaged Index for all public elementary and secondary schools, made available by the Ministère de l'Éducation et de l'enseignement supérieur (Ministry of Education and Post-Secondary Studies), provided a picture of the socio-economic background index (IMSE) for each school in our sample. Schools are ranked on a scale from 1 to 10, with a score of 1 representing the least disadvantaged schools and a score of 10 representing the most disadvantaged schools (Gouvernement du Québec. Ministère de l'Éducation et de l'Enseignement Supérieur, 2020; see **Table 2**).

Two fixed classrooms were in the same school having a score of 7. The school of the remaining fixed classroom had

TABLE 2 | Socio-economic environment index (IMSE) and decile rank for each school in the sample.

Classroom	Grade	School	Socio-economic environment index (IMSE)	Decile rank (IMSE)
Fixed group				
Classroom 1	5	School 1	9.52	7
Classroom 2	5	School 1	9.52	7
Classroom 3	6	School 2	9.15	6
Flexible group				
Classroom 4	5	School 3	3.79	2
Classroom 5	5	School 4	5.50	3
Classroom 6	6	School 5	8.77	6

TABLE 3 | Description of variables measured by BASC-3 (Reynolds and Kamphaus, 2015).

Internalizing problems	Inattention/ hyperactivity	School problems	Emotional symptoms	Personal adjustment
Atypicality; Locus of control; Social stress; Anxiety and depression; Sense of inadequacy	Attention problems; Hyperactivity	Attitude to school; Attitude to teachers	Social stress; Anxiety and Depression; Sense of inadequacy; Self-esteem; Self-reliance	Relations with parents; Interpersonal relations; Self-esteem; Self-reliance

a score of 6. Consequently, these classrooms were in the least advantaged schools in the CSS. As for the flexible classroom schools, one had the same disadvantage index score as the fixed classroom schools (6). However, the other two schools with flexible classrooms had scores of 2 and 3, respectively, meaning they were in the most advantaged schools in the CSS. Therefore, the schools were not equivalent overall with regard to their disadvantage index.

After receiving approval from the Institutional Human Research Ethics Committee (CIEREH) and authorization from the CSS, the school principals, and the teachers, we were able to meet with the students in class to present the study. Since participation in the project was voluntary, an invitation to participate and a consent form were distributed for parents to sign. Data collection took place during December 2019 (pre-pandemic). As such, the students were exposed to the research environment for a period of 4 months, from the beginning of the school year. The data collection procedure lasted, on average, about 40 min per student and was conducted under the supervision of a research assistant. Students were asked to complete two questionnaires (see "Students"): the Liddle and Carter (2015) questionnaire and the BASC-3 (Reynolds and Kamphaus, 2015). The Liddle and Carter (2015) questionnaire was distributed in paper format and took students 10 min on average to complete. The BASC-3 required 30 min on average to complete. Students completed the paper version, and their responses were transcribed using the Q-Global platform licensed by NCS Pearson, Inc. No participants withdrew during data collection.

Measuring Instruments and Data Collection Procedures

Teacher Matching

As mentioned above, the objective of the questionnaire was to match the three flexible classroom teachers with three fixed classroom teachers having a similar teacher profile in order to control for teacher effect. The self-report questionnaire consisted of twenty-four items divided into three categories. The *Teacher Self-Efficacy scale* consisted of ten items ($\alpha = 0.82$; Schwarzer, 1992; Bandura, 1997; Schwarzer and Hallum, 2008) in which teachers were asked to choose among four Likert scale responses (not at all true, only slightly true, moderately true, completely true). *Job accomplishment* was assessed using five items ($\alpha = 0.77$; Ho and Au, 2006). Teachers were asked to select the most appropriate response among five items (strongly disagree, somewhat disagree, neither disagree nor agree, somewhat agree, strongly agree). Finally, *Intention in the performance goal structure* was measured using nine items ($\alpha = 0.69$; Midgley et al., 2000). Teachers were asked to rate each statement on a Likert scale from 1 to 7, 1 being completely false and 7 being completely true. The questionnaire took an average of 10 min to complete.

Students

Students were asked to complete two self-reported questionnaires individually, the first measuring wellbeing at school using the Liddle and Carter (2015) questionnaire, and the second measuring mental health using the BASC-3 tool (Reynolds and Kamphaus, 2015), which reports on students' coping and behavioral functioning.

The Liddle and Carter (2015) questionnaire is a 12-item self-report questionnaire ($\alpha = 0.82$) using a Likert scale (never, not often, regularly, often, all the time). For example, students were asked to respond to the following statements: "I think good things will happen to me in my life"; "I get along with people"; and "I feel relaxed."[5]

The BASC-3 (Reynolds and Kamphaus, 2015) measurement tool examines the mental health of students by taking into account various aspects of personal adjustment and behavioral functioning. It consists of 137 items. The five composite scales (*internalizing problems, inattention/hyperactivity, school problems, emotional symptoms, personal adjustment*; $\alpha = 0.89$–0.95) grouped ten clinical scales (*anxiety, attention problems, attitude to school, attitude to teachers, atypicality, depression, hyperactivity, locus of control, sense of inadequacy, social stress*; $\alpha = 0.73$–0.86) and four adaptive scales (*relations with parents, interpersonal relations, self-esteem*, *self-reliance*; $\alpha = 0.75$–0.87; see **Table 3**). The questionnaire was divided into two parts. In the first part, students were to answer true or false for each statement. For example, students were asked to respond to the following statements: "I often do

[5]"Je pense que de bonnes choses vont m'arriver dans ma vie"; "Je m'entends bien avec les gens"; "Je me suis sentie détendue." [Our translation].

TABLE 4 | Basc-3 scale and composite score classification (Reynolds and Kamphaus, 2015).

T-score Range	Clinical Scales	Adaptive Scales
70 and above	Clinically significant	Very high
60–69	At risk	High
41–59	Average	Average
31–40	Low	At risk
30 and below	Very low	Clinically significant

things without thinking"; "I am not interested in school"; "I like who I am."[6] In the second part, students were to choose from four Likert scale items (never, sometimes, often, almost always). In this section, students were asked, among other things, to respond to the following statements: "I get along well with others"; "I am nervous"; and "I am a good listener."[7]

Data Analysis

To provide a portrait of each group, descriptive analyses were performed on the entire sample ($N = 107$) using the two variables of "age" and "gender." For each group, age was described by mean and standard deviation (SD). The "gender" variable was described by its frequency in each group (fixed and flexible classrooms).

As previously mentioned, the students' responses to the BASC-3 were transcribed using the Q-Global platform, thus providing an analysis of all students' scores. Scores were standardized, that is, raw scores were translated into T-scores and percentile scores (see **Table 4**).

Normality of distribution was verified beforehand using a normal probability plot with Henry's line, and homogeneity of variance was verified by a Levene test. For statistical analyses, STATA 15.1 software was used, with a significance level less than or equal to 0.05. A Student *t*-test was performed (independent variable "group" with two categories) to analyze the different variables measured by the instruments (quantitative dependent variables). Differential analyses were also conducted by gender (boys and girls separately). The results of the study are discussed in the next section.

RESULTS

To recall, the purpose of this exploratory study was to compare the wellbeing and mental health of Grade 5 and 6 students in differentiated classroom seating arrangements (fixed and flexible classrooms). Secondly, the study compared gender-specific wellbeing and mental health of Grade 5 and 6 students in differentiated classroom seating arrangements (fixed and flexible classrooms). The results are therefore presented in three parts:

1. Mental health indicator results by group (fixed and flexible) for both genders.
2. Mental health indicator scores by group (fixed and flexible) for boys.
3. Mental health indicator scores by group (fixed and flexible) for girls.

Mental Health Indicator Results by Group (Fixed and Flexible) for Both Genders

Table 5 presents the mental health indicator scores by group (fixed and flexible) for both genders. For each mental health indicator, the mean (M), standard deviation (SD), minimum, and maximum obtained by students in each group, as well as statistical significance (*p*-value), are presented.

For the "wellbeing" variable, the difference was not statistically significant ($p = 0.197$). Students in the flexible group had a lower mean wellbeing score of 57.1 points (SD = 8.57) compared to a mean wellbeing score of 54.8 points (SD = 9.57) for students in the fixed group.

For the "internalizing problems" variable, students in the fixed group had a slightly higher mean score ($M = 53.1$; SD = 11.7) than students in the flexible group ($M = 52.4$; SD = 10.5). Although this difference was not statistically significant ($p = 0.740$), the scores suggest that students in the flexible group exhibited less atypicality, social stress, anxiety, and depression. They appeared to have a better locus of control and a lower sense of inadequacy compared to students in the fixed group.

For the "inattention/hyperactivity" variable, students in the fixed group had a mean score of 50.9 points (SD = 9.30), while students in the flexible group had a mean score of 51.9 points (SD = 11.0). Thus, students in the flexible group reported relatively more attention problems with or without hyperactivity than students in the fixed group. However, this difference was not statistically significant ($p = 0.618$).

TABLE 5 | Mental health indicator results by group (fixed and flexible) for both genders.

	Indicator	Mean (SD)	Minimum–Maximum	*p*-value
Fixed group	Wellbeing (*n* = 51)	54.8 (9.57)	33.5–72.0	0.197
	Mental health (*n* = 51)			
	Internalizing problems (*n* = 51)	53.1 (11.7)	35.0–79.0	0.740
	Inattention/hyperactivity (*n* = 51)	50.9 (9.30)	35.0–73.0	0.618
	School problems (*n* = 51)	49.3 (9.19)	37.0–74.0	0.426
	Emotional symptoms (*n* = 51)	51.5 (11.0)	35.0–77.0	0.901
	Personal adjustment (*n* = 51)	52.3 (7.66)	29.0–63.0	0.963
Flexible group	Wellbeing (*n* = 56)	57.1 (8.57)	39.0–74.0	0.197
	Mental health (*n* = 56)			
	Internalizing problems (*n* = 56)	52.4 (10.5)	36.0–90.0	0.740
	Inattention/hyperactivity (*n* = 56)	51.9 (11.0)	34.0–75.0	0.618
	School problems (*n* = 56)	48.0 (8.04)	37.0–74.0	0.426
	Emotional symptoms (*n* = 56)	51.3 (10.6)	36.0–92.0	0.901
	Personal adjustment (*n* = 56)	51.5 (9.46)	23.0–63.0	0.963

[6]"Je fais souvent des choses sans réfléchir"; "L'école ne m'intéresse pas"; "J'aime qui je suis" [Our translation].

[7]"Je m'entends bien avec les autres"; "Je suis nerveux"; "Je sais bien écouter." [Our translation].

TABLE 6 | Mental health indicator results by group (fixed and flexible) for boys.

	Indicator	Mean (SD)	Minimum–Maximum	*p*-value
Fixed group	Wellbeing (*n* = 27)	55.3 (9.84)	33.5–72.0	0.492
	Mental health (*n* = 27)			
	Internalizing problems (*n* = 27)	50.2 (9.86)*	35.0–79.0	0.016
	Inattention/hyperactivity (*n* = 27)	50.7 (7.99)**	36.0–67.0	0.010
	School problems (*n* = 27)	53.1 (9.25)	39.0–74.0	0.454
	Emotional symptoms (*n* = 27)	48.9 (8.69)*	35.0–71.0	0.015
	Personal adjustment (*n* = 27)	52.9 (6.30)	41.0–63.0	0.064
Flexible group	Wellbeing (*n* = 30)	53.6 (9.02)	39.0–74.0	0.492
	Mental health (*n* = 30)			
	Internalizing problems (*n* = 30)	56.8 (10.3)*	38.0–90.0	0.016
	Inattention/hyperactivity (*n* = 30)	57.2 (10.3)**	37.0–75.0	0.010
	School problems (*n* = 30)	51.3 (8.51)	39.0–74.0	0.454
	Emotional symptoms (*n* = 30)	55.6 (11.1)*	38.0–92.0	0.015
	Personal adjustment (*n* = 30)	47.2 (10.4)	23.0–63.0	0.064

*$p \leq 0.050$; **$p \leq 0.010$.

TABLE 7 | Mental health indicator results by group (fixed and flexible) for girls.

	Indicator	Mean (SD)	Minimum–Maximum	*p*-value
Fixed group	Wellbeing (*n* = 24)	54.2 (9.44)**	39.0–72.0	0.003
	Mental health (*n* = 24)			
	Internalizing problems (*n* = 24)	56.3 (12.9)**	38.0–75.0	0.004
	Inattention/hyperactivity (*n* = 24)	51.2 (10.8)*	35.0–73.0	0.050
	School problems (*n* = 24)	45.1 (7.19)	37.0–60.0	0.607
	Emotional symptoms (*n* = 24)	54.5 (12.7)**	38.0–77.0	0.007
	Personal adjustment (*n* = 24)	51.7 (9.04)	29.0–62.0	0.073
Flexible group	Wellbeing (*n* = 26)	61.1 (5.96)**	50.0–70.0	0.003
	Mental health (*n* = 26)			
	Internalizing problems (*n* = 26)	47.0 (8.01)**	36.0–68.0	0.004
	Inattention/hyperactivity (*n* = 26)	45.8 (8.33)*	34.0–68.0	0.050
	School problems (*n* = 26)	44.1 (5.46)	37.0–54.0	0.607
	Emotional symptoms (*n* = 26)	46.3 (7.58)**	36.0–77.0	0.007
	Personal adjustment (*n* = 26)	56.3 (4.87)	45.0–63.0	0.073

*$p \leq 0.050$; **$p \leq 0.010$.

As for the "school problems" variable, which reflects student attitudes to school and teachers, the mean scores were 49.3 points (SD = 9.19) for students in the fixed group and 48.0 points (SD = 8.04) for students in the flexible group. However, this difference was not statistically significant ($p = 0.426$).

For the "emotional symptoms" variable, there were no statistically significant differences ($p = 0.901$) between the two groups. The mean scores were 51.5 points (SD = 11.0) for students in the fixed group and 51.3 points (SD = 10.6) for students in the flexible group. The students therefore had comparable mean scores for social stress, anxiety, depression, and sense of inadequacy, as well as self-esteem and self-reliance.

The difference was not statistically significant ($p = 0.963$) for the "personal adjustment" variable. Students in the fixed group had a score of 52.3 points (SD = 7.66), while students in the flexible group had a mean score of 51.5 points (SD = 9.46). This suggests that students in the flexible group had relatively better parent–child and interpersonal relations, higher self-esteem, and higher levels of self-reliance.

Mental Health Indicator Results by Group (Fixed and Flexible) for Boys

Table 6 presents the mean scores [standard deviation (SD), minimum, and maximum] for boys' mental health indicators by group (fixed and flexible), as well as the statistical significance (*p*-value) of each indicator.

Regarding the "wellbeing" variable, although the difference was not statistically significant ($p = 0.492$), boys in the fixed group had a higher mean score ($M = 55.3$; SD = 9.84) than those in the flexible group ($M = 53.6$; SD = 9.02).

The difference was statistically significant ($p = 0.016$) for the "internalizing problems" variable. In the flexible group, boys had a higher mean score ($M = 56.8$; SD = 10.3) than boys in the fixed group ($M = 50.2$; SD = 9.86). As a result, boys in the fixed group reported less atypicality, social stress, anxiety, and depression, and had a better locus of control and a lower sense of inadequacy compared to boys in the flexible group.

The mean scores for the "inattention/hyperactivity" variable were 50.7 points (SD = 7.99) for boys in the fixed group and 57.2 points (SD = 10.3) for boys in the flexible group. Therefore, boys in the fixed group presented less attention problems with or without hyperactivity than boys in the flexible group. The difference was statistically significant ($p = 0.010$).

As for the "school problems" variable, boys in the flexible group ($M = 51.3$; SD = 8.51) had relatively better attitudes to school and teachers compared to boys in the fixed group ($M = 53.1$; SD = 9.25). However, this difference was not statistically significant ($p = 0.454$).

For the "emotional symptoms" variable, the difference between the two groups was statistically significant ($p = 0.015$). The mean scores were 48.9 points (SD = 8.69) for boys in the fixed group and 55.6 points (SD = 11.1) for boys in the flexible group. Thus, boys in the flexible group had more social stress, anxiety, and depression. They also had a greater sense of inadequacy, lower self-esteem, and lower levels of self-reliance than students in the fixed group.

Regarding the "personal adjustment" variable, although the difference was not statistically significant ($p = 0.064$), boys in the fixed group had a higher mean score (M = 52.9; SD = 6.30) than boys in the flexible group ($M = 47.2$; SD = 10.4). Boys in the fixed group tended to have better interpersonal and family relationships, self-esteem, and self-reliance compared to boys in the flexible group, although the observed difference did not reach statistical significance.

Mental Health Indicator Results by Group (Fixed and Flexible) for Girls

Table 7 presents the scores (mean, standard deviation [SD], minimum, and maximum) for mental health indicators by group (fixed and flexible) for girls, as well as the statistical significance (*p*-value) of each indicator.

For the "wellbeing" variable, girls in the flexible group had a higher mean score ($M = 61.1$; SD = 5.96) than girls in the fixed group ($M = 54.2$; SD = 9.44). This difference between the two groups of girls was statistically significant ($p = 0.003$).

There was a statistically significant difference between the two groups of girls for the "internalizing problems" variable ($p = 0.004$). In the fixed classrooms, girls had a higher mean score ($M = 56.3$; SD = 12.9) than girls in the flexible classes ($M = 47.0$; SD = 8.01). Based on the results, girls in the fixed classrooms showed more internalizing problems than girls in the flexible classrooms. Thus, girls in the fixed group showed more atypicality, had greater social stress, anxiety, depression, and sense of inadequacy, and had poorer locus of control compared to girls in the flexible group.

For the "inattention/hyperactivity" variable, the difference was statistically significant ($p = 0.050$). Girls in the flexible group ($M = 45.8$; SD = 8.33) had fewer attention problems with or without hyperactivity than girls in the fixed group ($M = 51.2$; SD = 10.8).

Regarding school problems, girls in the fixed group were found to have a slightly higher mean score ($M = 45.1$; SD = 7.19) compared to girls in the flexible group ($M = 44.1$; SD = 5.46). Although the difference did not reach statistical significance ($p = 0.607$), the mean score for school problems for girls in the flexible group was lower (−1 point).

For the "emotional symptoms" variable, girls in the flexible group ($M = 46.3$; SD = 7.58) had a lower mean score compared to girls in the fixed group ($M = 54.5$; SD = 12.7) and thus had less social stress, anxiety, depression, had a lower sense of inadequacy, and had higher self-esteem and self-reliance than girls in the fixed group. This difference was statistically significant ($p = 0.007$).

The mean scores for the "personal adjustment" variable were 51.7 points (SD = 9.04) for girls in the fixed group and 56.3 points (SD = 4.87) for girls in the flexible group. Although this difference was not statistically significant ($p = 0.073$), girls in the fixed group tended to have poorer interpersonal and family relationships, self-esteem, and self-reliance.

DISCUSSION

The results presented in the previous section are discussed below. First, we will answer the research question. The results will then be discussed in light of the available scientific literature. Finally, the study's contributions, limitations, and prospects for research will be discussed.

Summary of Results and Answer to the Research Question

With regard to the above results, the comparative analysis of the two groups (fixed and flexible) for both genders combined did not reveal any statistically significant difference for the various variables. However, when the two groups were analyzed by gender, statistically significant differences were found between the two groups (fixed and flexible). The mean scores for boys showed statistically significant differences for the following variables: (1) internalizing problems, (2) inattention and hyperactivity, and (3) emotional symptoms. Thus, boys in the fixed group had significantly fewer internalizing problems, attention problems with or without hyperactivity, and emotional symptoms than boys in the flexible group. For girls, the statistically significant variables were (1) wellbeing, (2) internalizing problems, (3) inattention and hyperactivity, and (4) emotional symptoms. Unlike the boys, girls in the flexible group reported greater wellbeing and fewer internalizing problems, attention problems with or without hyperactivity, and emotional symptoms. Based on the results of our gender-differentiated analysis, it appears that classroom seating arrangement influenced the wellbeing and mental health of elementary students at school. Based on the data, boys had a greater sense of wellbeing and mental health in fixed classrooms. In contrast, among the girls, the classroom seating arrangement most conducive to their wellbeing and mental health, according to these results, was flexible seating. Thus, flexible seating seemed to be a real challenge for some students and a real asset for others, which we will now discuss.

Flexible Seating: Advantages and Limitations

As noted above, previous research on flexible classroom seating has reported that this type of arrangement helps meet students' needs (Comaianni, 2017; Havig, 2017; Limpert, 2017; Erz, 2018; Legout, 2018; Schoolcraft, 2018; Schrage, 2018; Sorrell, 2019; Vallée, 2019) and encourages the development of skills, such as self-reliance, self-regulation, and problem-solving (Doyon, 2018; Erz, 2018; Laquerre, 2018; Legout, 2018). Although flexible seating is intended to be student-centered and needs-based, our results indicate that this type of arrangement can be detrimental to the wellbeing and mental health of some students. In the flexible classroom, students no longer have a place assigned to them (Legout, 2018). They move about freely and choose the seat that best suits the task at hand. As a result, the flexible classroom requires students to apply more skills, such as self-control, problem-solving, self-reliance, cooperation, and soft skills, such as working together, and so on. Flexible seating may therefore require students to initially have good coping strategies.

Surveys in Quebec have reported that girls perform better in problem-solving and self-control skills, among other things (Direction régionale de santé publique de Montréal, 2018; Institut de la statistique du Québec, 2018). However, problem-solving is a critical coping skill in the flexible classroom since students are required to make strategic choices throughout the day (Dornfeld, 2016; Havig, 2017; Limpert, 2017; Del'Homme, 2018; Laquerre, 2018; Legout, 2018; Tiennot, 2019; Vallée, 2019). As for the skill of self-control, it allows students to self-regulate more readily (Félouzis, 1993; Bouchard et al., 2006; Besnard et al., 2016) and makes it easier to adapt to the norms and expectations of the school (Commissariat général à la stratégie et à la prospective, 2014; Esperbès-Pistre et al., 2015). Self-control is especially important in flexible

seating to be able to exercise self-reliance and cooperation in a classroom where all the furniture is available to the students. In addition, it appears that girls tend to develop more pro-social behaviors conducive to cooperation in the classroom (Félouzis, 1993; Bouchard et al., 2006; Ruel, 2010; Besnard et al., 2016), a key aspect of the flexible classroom (Del'Homme, 2018). Thus, girls may have an easier time adapting and engaging in the flexible classroom, which would explain why girls in the flexible classrooms had a higher sense of wellbeing and mental health, as measured. More broadly, flexible seating, through the practices, behaviors, and attitudes it encourages, may benefit students who initially have good coping strategies, may enhance their sense of control and may help meet their needs (self-reliance, socialization, and so on). This type of classroom arrangement may therefore be conducive to their wellbeing and mental health.

However, students who have difficulty adapting and behaving in a way that is conducive to the task at hand may be challenged by the flexible classroom. Many studies have reported that boys are more affected by behavioral and learning problems (Walker and Berthelsen, 2007; Childs and McKay, 2010; Besnard et al., 2016), which may affect their concentration, on-task behavior (Félouzis, 1991, 1993; Bouchard et al., 2006; Walker and Berthelsen, 2007; Ruel, 2010; Girardin, 2012; Gilles, 2018), and coping skills. Some studies have noted that freedom of choice and movement can be challenging for students who need a framework and routine (Legout, 2018; Schoolcraft, 2018; Vallée, 2019). For these students, fixed seating appears to be beneficial to their wellbeing and mental health. This may be because having an assigned desk, in other words, a space of their own, is reassuring (Legout, 2018) and reinforces their sense of control. Moreover, fixed seating provides a framework that may be more appropriate for these students (Legout, 2018; Vallée, 2019). Thus, it is not so much flexible furniture *per se* that may explain why these students have a lower sense of wellbeing and mental health, but how the flexible classroom itself functions (less controlling environment, undefined personal space, and so on; Havig, 2017; Legout, 2018; Vallée, 2019). Nevertheless, flexible seating is not to be ruled out for students with coping difficulties, but it does require teachers to provide alternatives and more ongoing support for these students.

Contributions of the Study

These results add to current knowledge in the field of educational research. Because flexible seating is a recent phenomenon, few studies have been conducted on the topic (Havig, 2017; Laquerre, 2018; Vallée, 2019) and little is known about the influence of flexible seating on student wellbeing and mental health. Furthermore, few studies have compared the two types of seating arrangements (fixed versus flexible classrooms), and those that do rarely conduct gender-differentiated analyses.

On a practical level, this study provides additional guidance for teachers. It invites teachers to better anticipate the potential limitations of flexible seating to better prepare students for change. Indeed, regular support by teachers for students who need to develop coping strategies would seem vital.

Limitations of the Study

Some of the limitations of our study relate to our sample. First, the small sample size ($N = 107$) does not allow drawing generalizable conclusions from our results. Indeed, this study was exploratory and intended to generate hypotheses and research questions in a new field of research. Another limitation of our sample lies in the socio-economic background indices (IMSE). Although our study only considered the socio-economic factor of school in his neighborhood, it did not consider the socio-economic factor of each family. This study was part of an exploratory process at the start, and we did not plan to collect this data directly from the parents of students. Thus, concerning the distribution of the presence of psychological difficulties (e.g., internalized and externalized problems and ADHD) in the two groups, there could be misleading reading of the results.

Furthermore, recall that the schools in the fixed group were among the least advantaged schools in the CSS, while two schools in the flexible group were among the most advantaged schools. However, previous research indicates that difficult socio-economic conditions in the home environment are associated with lower sense of wellbeing, mental health, and academic achievement (Ayotte et al., 2009; Riberdy et al., 2013; Couture, 2019). According to Riberdy et al. (2013), youth from disadvantaged backgrounds are more likely to be diagnosed with a mental health problem and report a perceived mental health problem. As a result, a higher prevalence of behavioral problems (hyperactivity, internalizing, and externalizing problems) is observed in these youth (Ayotte et al., 2009; Riberdy et al., 2013; Kettani et al., 2017; Couture, 2019). Moreover, according to Childs and McKay (2010), boys appear to be more susceptible to the effects of a low socio-economic background. Thus, the poorer mental health of boys in the flexible group compared to boys in the fixed group does not appear to be explained by socio-economic background.

A number of limitations of the study relate to methodology. To better understand the results, it would have been useful to use a mixed design and incorporate qualitative data through individual or group interviews. Also, our starting methodology had to have two measurement times, which would have been optimal for answering our research questions. This constitutes a significant limitation to our study and to the interpretation of the results concerning the differences between our groups. Moreover, it would have been useful to do a second measurement at the end of the year (outside of the pandemic context) to compare the two groups over the school year and to see if there were any changes in mental health indicators. In this sense, an important limit of the results indicates that the independent variable (type of arrangement) has a significant effect, not so much on the processes of adaptation/functioning in the classroom, but on indicators of general functioning (internalizing problems; inattention/hyperactivity; emotional symptoms) which can probably be interpreted as previous aspects, not attributable simply to the arrangement of the class.

Another limitation is that not all the data collected were independent from one another. Indeed, the analyses conducted (*t*-tests) were based on the premise of non-independence of

the data. This limitation could have been circumvented by introducing group affiliation as a covariate in the analyses. Furthermore, the gender-differentiated analysis used separate *t*-tests, but this distinction required first validating that significant interaction emerged *a priori*. This would have required conducting a MANOVA predicting mental health indicators and wellbeing and including gender and group as inter-subject variables. Were this interaction significant, separate *t*-tests would have been indicated. However, given the gender significance of the results, there is little doubt that this interaction was significant. The other limitation of our study concerns the limited literature on the topic. Indeed, our results could not be documented and supported by other studies that conducted gender-differentiated analyses. The final limitation of our study is that we did not consider students with special needs. A study should focus on the inclusion of these students in the context of flexible seating classroom.

Therefore, the above considerations need to be confirmed. Ultimately, this study had an exploratory intention. In addition, the protocol had to be modified because of COVID-19, a measurement time could not be completed. As we cannot redo the study, we can only add limits to the discussion and place the study in an exploratory context of research on a seed grant model in a new field.

Prospects for Research

Our findings suggest the need for further studies on the topic. Indeed, the results of this study provide initial data on the influence of classroom seating arrangement on student wellbeing and mental health at school.

In view of the differences found between boys and girls, it would seem vital to make gender-differentiated analyses routine in scientific research related to human health or behavior (Tannenbaum et al., 2019). In the future, it would be relevant for studies conducting qualitative analyses to gather student and teacher perceptions to have a better understanding of school design-related factors influencing student mental health and wellbeing.

Also, it should be emphasized that the pedagogical methods within the classrooms affect a positive modification of the learning processes in children compared to the fixed seating classroom. In fact, in this sense, in a future study, pedagogy and learning processes should also be investigated in addition to wellbeing.

Finally, it would be very interesting in future studies in this field to favor a comparison between different cultures with an intercultural perspective. Thus, flexible classrooms could promote inclusive processes for children with special educational needs.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Human Research Ethics Committee (CIEREH) of the Université du Québec à Montréal (UQAM) and bears the following ethics certificate number: 3761_e_2019. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This study was funded by the Social Sciences and Humanities Research Council of Canada (SSHRC) under an Insight Development Grant (reference number: 430-2019-00875).

ACKNOWLEDGMENTS

We would like to thank the Montreal South Shore School Services Centre for their interest and assistance in this project.

REFERENCES

Abbasi, N. (2013). Organisation de l'espace scolaire et formation de l'identité chez les adolescents: Une analyse de quatre établissements secondaires en Australie. *Revue internationale d'éducation de Sèvres* 64, 133–145. doi: 10.4000/ries.3626

Allen, C. (2018). Flexible Seating: Effects of Student Seating Type Choice in the Classroom. Master's Thesis. Macomb (IL): Western Illinois University.

Amoly, E., Dadvand, P., Forns, J., López-Vicente, M., Basagaña, X., Julvez, J., et al. (2014). Green and blue spaces and behavioral development in Barcelona schoolchildren: The BREATHE project. *Environ. Health Perspect.* 122, 1351–1358. doi: 10.1289/ehp.1408215

Antoine, P., Poinsot, R., and Congard, A. (2007). Évaluer le bien-être subjectif: la place des émotions dans les psychothérapies positives. *Journal de Thérapie Comportementale et Cognitive.* 17, 170–180. doi: 10.1016/S1155-1704(07)78392-1

Ayotte, V., Fournier, M., and Riberdy, H. (2009). *La détresse psychologique des enfants et des adolescents montréalais… l'expression de différentes réalités? Enquête sur le bien-être des jeunes Montréalais*. Montréal: Direction de santé publique, Agence de la santé et des services sociaux de Montréal.

Aziz, N. F., Noor, N. N. M., Salleh, N. M., and Said, I. (2017). A review on the significance of school grounds design on children's performances and well-being. *Int. J. Env. Soc. Space* 5, 93–100.

Bacro, F., Guimard, P., Florin, A., Ferrière, S., and Gaudonville, T. (2017). Bien-être perçu, performances scolaires et qualité de vie des enfants à l'école et au collège: étude longitudinale. *Enfance* 2017, 61–80. doi: 10.4074/S0013754517001057

Bandura, A. (1997). *Self–Efficacy: The Exercise of Control*. New York: Freeman.

Barrett, P., Davies, F., Zhang, Y., and Barrett, L. (2017). The holistic impact of classroom spaces on learning in specific subjects. *Environ. Behav.* 49, 425–451. doi: 10.1177/0013916516648735

Barrett, P., Zhang, Y., Davies, F., and Barrett, L. (2015). *Clever Classrooms: Summary Report of the HEAD Project*. Salford: University of Salford.

Besnard, T., Lemelin, J. P., and Houle, A.-A. (2016). Différences dans le niveau de développement des garçons et des filles et efficacité différenciée des interventions selon le genre à la période préscolaire. Université de Sherbrooke: Groupe de recherche sur les inadaptations de l'enfance (GRISE). Available at: http://www.grise.ca/images/docs/Rapport_Final_Besnard_Lemelin_MSSS_2mai2016.pdf (Accessed April 30, 2022).

Bluteau, J., Aubenas, S., Dufour, F., and Carrier, L. (2019). La classe flexible est-elle un passage de mode? *La Foucade.* 20, 19–20.

Blyth, A. (2013). Perspectives pour les futurs espaces scolaires. *Revue internationale d'éducation de Sèvres.* 64, 53–64. doi: 10.4000/ries.3606

Bouchard, C., Cloutier, R., and Gravel, F. (2006). Différences garçons-filles en matière de prosocialité. *Enfance* 58, 377–393. doi: 10.3917/enf.584.0377

Boudreault, M. -A. (2017). Contrer la sédentarité des élèves avec le « flexible seating ». Réseau d'information pour la réussite éducative. Available at: http://rire.ctreq.qc.ca/2017/06/flexible-seating/ (Accessed April 30, 2022).

Bronfenbrenner, U., and Morris, P. A. (1998). "The ecology of developmental processes," in *Handbook of Child Psychology: Theoretical Models of Human Development.* eds. D. W. Damon and R. -M. Lerner (New York: John Wiley), 993–1028.

Broto, C. (2013). Architecture pour l'éducation. A. Aussel et L. Barros, trad. Links Books.

Carignan, M. A. (2018). *Les écoles qu'il nous faut*. Montréal: Éditions MultiMondes.

Cheryan, S., Ziegler, S. A., Plaut, V. C., and Meltzoff, A. N. (2014). Designing classrooms to maximize student achievement. *Policy Insights Behav. Brain Sci.* 1, 4–12. doi: 10.1177/2372732214548677

Childs, G., and McKay, M. (2010). Boys starting school disadvantaged: implications from teachers' ratings of behaviour and achievement in the first two years. *Br. J. Educ. Psychol.* 71, 303–314. doi: 10.1348/000709901158532

Comaianni, K. (2017). What is The impact of flexible seating on students in grade one at one elementary school. Master's Thesis. San Marcos (CA): California State University san Marcos.

Commissariat général à la stratégie et à la prospective. (2014). *Lutter contre les stéréotypes filles-garçons: un jeu d'égalité et de mixité dès l'enfance*. Paris: l'auteur.

Conseil Supérieur de l'Éducation. (2020). *Le bien-être de l'enfant à l'école: faisons nos devoirs*. Québec: Gouvernement du Québec.

Couture, H. (2019). *La santé mentale des enfants et des adolescents: Données statistiques et enquêtes recensées, Études et recherches*. Québec: Conseil supérieur de l'Éducation.

Del'Homme, M. (2018). Mise en place d'une classe flexible: quels impacts sur les compétences des élèves? Master's Thesis. Nantes (FR): Université de Nantes.

Delzer, K. (2015). Reimagining classrooms: teachers as learners and students as leaders conference presentation TEDx talks 2015. Available at: https://www.youtube.com/watch?v=w6vVXmwYvgs (Accessed April 30, 2022).

Direction régionale de santé publique de Montréal (2018). *Enquête Topo: portrait des jeunes montréalais de 6e année*. Québec: l'auteur.

Doré, I., and Caron, J. (2017). Santé mentale: concepts, mesures et déterminants. *Sante Ment. Que.* 42, 125–145. doi: 10.7202/1040247ar

Dornfeld, K. (2016). Flexible seating and student-centered classroom redesign: how a visit to Starbucks inspired a change to flexible seating and then some changes in teaching philosophy. Edutopia. Available at: https://www.edutopia.org/blog/flexible-seating-student-centered-classroom-kayla-delzer (Accessed April 22, 2016).

Doyon, K. (2018). L'applicabilité des principes de positionnement de l'aménagement flexible par des élèves de 5ème et 6ème années du primaire. Master's Thesis. Trois-Rivières (QC): Université du Québec à Trois-Rivières.

Eccles, J.-S., and Roeser, R.-W. (2011). "School and community influences on human development," in *Developmental Science: An Advanced Textbook*. eds. M.-H. Borstein and M.-E. Lamb (New York: Psychology Press), 571–643.

Erz, S. L. (2018). Impact and Implications of the Flexible Learning Environment in the at-Risk Secondary Classroom. Master's Thesis. Minot (ND): Minot State University.

Esperbès-Pistre, S., Bergonnier-Dupuy, G., and Cazenave-Tapie, P. (2015). Le stress scolaire au collège et au lycée: différences entre filles et garçons. *Éducation et francophonie.* 43, 87–112. doi: 10.7202/1034487ar

Espinosa, G., and Rousseau, N. (2018). "Le bien-être à l'école et l'apport de la psychopathologie," in *Le bien-être à l'école: enjeux et stratégies gagnantes* (ed.) N. Rousseau et G. Espinosa (dir.) (Qubec: Presses de l'Université du Québec), 1–11. doi: 10.1787/9789264292505-fr

Farmer, T. W., Lines, M. M., and Hamm, J. V. (2011). Revealing the invisible hand: The role of teachers in children's peer experiences. *J. Appl. Dev. Psychol.* 32, 247–256. doi: 10.1016/j.appdev.2011.04.006

Félouzis, G. (1991). Comportements de chahut et performances scolaires des filles et des garçons au collège. *Enquête* 6, 1–13. doi: 10.4000/enquete.140

Félouzis, G. (1993). Interactions en classe et réussite scolaire: une analyse différenciée filles-garçons. *Rev. Fr. Sociol.* 34, 199–222. doi: 10.2307/3322488

Ferrière, S., Bacro, F., Florin, A., and Guimard, P. (2016). Le bien-être en contexte scolaire: intérêt d'une approche par triangulation méthodologique. *Les Cahiers Internationaux de Psychologie Sociale.* Numéro 111, 341–365. doi: 10.3917/cips.111.0341

Florin, A., and Guimard, P. (2017). La qualité de vie à l'école: Comment l'école peut-elle proposer un cadre de vie favorable à la réussite et au bien-être des élèves. Paris: Conseil National d'évaluation du système scolaire (CNESCO).

Fouquet-Chauprade, B. (2013). L'ethnicité au collège: bien-être et effet de contexte. *Sociologie* 4, 431–449. doi: 10.3917/socio.044.0431

Gest, S. D., and Rodkin, P. C. (2011). Teaching practices and elementary classroom peer ecologies. *J. Appl. Dev. Psychol.* 32, 288–296. doi: 10.1016/j.appdev.2011.02.004

Gilles, E. (2018). Une expérience genrée des espaces du quotidien à l'adolescence: le cas des filles et des garçons de 4ème dans le Calvados et la Manche. Dissertation. Caen (FR): Université de Caen.

Girardin, D. (2012). L'oral en classe: garçons et filles, quelles différences? Master's Thesis. Genève (CHE): Université de Genève.

Gouvernement du Québec Ministère de l'Éducation et de l'Enseignement Supérieur (2020). Indices de défavorisation des écoles publiques: 2019-2020. Québec: l'auteur Available at: http://www.education.gouv.qc.ca/fileadmin/site_web/documents/PSG/statistiques_info_decisionnelle/Indices-defavorisation-2019-2020.pdf (Accessed June 11, 2020).

Guimard, P., Bacro, F., Ferrière, S., Florin, A., Gaudonville, T., and Ngo, U.-T. (2015). Le bien-être des élèves à l'école et au collège: validation d'une échelle multidimensionnelle, analyses descriptives et différentielles. *Éducation et formations.* 88, 163–184.

Havig, J. S. (2017). Advantages and disadvantages of flexible seating. Dissertation. Minot (ND): Minot State University.

Heaney, C., and Israel, B. (2008). "Social networks and social support," in *Health Behavior and Health Education: Theory, Research, and Practice. 4th Edn.* eds. K. Glanz, B.-K. Rimer and K. Viswanath (United States: Jossey-bass), 189–210.

Ho, C. L., and Au, W. T. (2006). Teaching satisfaction scale. Measuring job satisfaction of teachers. *Educ. Psychol. Meas.* 66, 172–185. doi: 10.1177/0013164405278573

Huynh, Q., Craig, W., Janssen, I., and Pickett, W. (2013). Exposure to public natural space as a protective factor for emotional well-being among young people in Canada. *BMC Public Health* 13, 1–14. doi: 10.1186/1471-2458-13-407

Institut de la statistique du Québec (2018). *Enquête québécoise sur la santé des jeunes du secondaire 2016–2017*. Québec: Gouvernement du Québec.

Jodelet, D. (2015). "Processus de mise en sens de l'espace et pratiques sociales." in *Digital Proceedings of the Symposium with International Participation: Places for Learning Experiences. Think, Make, Change*. eds. D. Germanos and M. Liapi; Grèce: Greek National Documentation Centre, 66–77.

Joing, I., Vors, O., Llena, C., and Potdevin, F. (2018). Se sentir bien dans chacun des lieux de l'espace scolaire au collège: Le rôle de l'autonomie, de l'appartenance sociale, de la perception du lieu et du sentiment de sécurité. *Spirale-Revue de recherches en éducation.* 61, 19–40. doi: 10.3917/spir.hs2.0019

Kettani, M., Zaouche-Gaudron, C., Lacharité, C., Dubeau, D., and Clément, M. È. (2017). Expérience paternelle et problèmes intériorisés de jeunes enfants en situation de précarité: le point de vue des pères. *Enfances, familles, générations.* 26:4ar. doi: 10.7202/1041064ar

Keymeulen, R., Henry, J., and Longlez, A. (2020). *La classe flexible: je me lance !: Cycles 1,2 et 3 et secondaire*. Paris: De Boeck Supérieur.

Konu, A., and Rimpelä, M. (2002). Well-being in schools: a conceptual model. *Health Promot. Int.* 17, 79–87. doi: 10.1093/heapro/17.1.79

Kruger, D.-J., Reischl, T.-M., and Gee, G.-C. (2007). Neighborhood social conditions mediate the association between physical deterioration and mental health. *Am. J. Community Psychol.* 40, 261–271. doi: 10.1007/s10464-007-9139-7

Laguardia, J.-G., and Ryan, R.-M. (2000). Buts personnels, besoin psychologiques fondamentaux et bien-être: théorie de l'autodétermination et applications. *Revue québécoise de psychologie.* 21, 281–304.

Laquerre, G. (2018). Nouvelle ère pour l'environnement d'apprentissage chez les élèves au primaire: Les classes flexibles, plus qu'un simple aménagement, un cheminement réflexif. Master's Thesis. Trois-Rivières (QC): Université du Québec à Trois-Rivières.

Legout, M. (2018). La classe flexible au service de la motivation, de la dynamique de groupe et de la différenciation. Master's Thesis. Bordeaux (FR): Université de Bordeaux.

Liddle, I., and Carter, G. F. (2015). Emotional and psychological well-being in children: the development and validation of the Stirling Children's well-being scale. *Educ. Psychol. Pract.* 31, 174–185. doi: 10.1080/02667363.2015.1008409

Limpert, S. M. (2017). A qualitative study of learning spaces at Midwest elementary school and its relationships to student attitudes about reading. Dissertation. St. Charles (MO): Lindenwood university.

Mazalto, M. (2017). *Concevoir des espaces scolaires pour le bien-être et la réussite*. Paris: L'Harmattan.

Mazalto, M., and Paltrinieri, L. (2013). Introduction: espaces scolaires et projets éducatifs. *Revue internationale d'éducation de Sèvres.* 64, 31–40. doi: 10.4000/ries.3592

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., and Freeman, K. E., … Urdan, T., (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*, Ann Arbor, MI: University of Michigan.

Organisation de coopération et de développement économiques (2001). Architecture et apprentissages: 55 établissements d'enseignement exemplaires. Paris: l'auteur. Récupéré de Available at: https://www.oecd-ilibrary.org/education/architecture-et-apprentissage_9789264292505-fr (Accessed May 14, 2019).

Organisation de coopération et de développement économiques (OCDE) (2011). Designing for education: compendium of exemplary educational facilities. Paris: l'auteur. Récupéré de

Pianta, R.-C., LaParo, K., and Hamre, B. (2008). *Classroom Assessment Scoring System Pre-K Manual*. Baltimore: Brookes Publishing.

Piché, G., Cournoyer, M., Bergeron, L., Clément, M.-È., and Smolla, N. (2017). Épidémiologie des troubles dépressifs et anxieux chez les enfants et les adolescents québécois. *Sante Ment. Que.* 42, 19–42. doi: 10.7202/1040242ar

Reynolds, C. R., and Kamphaus, R. W. (2015). *BASC-3 Behavior Assessment System for Children: Self-Report of Personality Child 8–11* (*3rd Edn.*). Toronto: Pearson Assessment Canada.

Riberdy, H., Tétreault, K., and Desrosiers, H. (2013). La santé physique et mentale des enfants: une étude des prévalences cumulatives. *Institut de la statistique du Québec.* 26, 1–28.

Rousseau, N. (2012). *Modèle dynamique de changement accompagné en contexte scolaire: pour le bien-être et la réussite de tous*. Québec: Presses de l'Université du Québec.

Ruel, S. (2010). L'espace classe: Structure de gestion de la construction culturelle des sexes pour les enfants de l'école élémentaire. *Agora débats/jeunesses.* 2, 55–66. doi: 10.3917/agora.055.0055

Schoolcraft, N. (2018). Student and teacher perceptions of stability balls as alternative seating in a first-grade classroom. *J. Teach. Act. Res.* 4, 90–111.

Schrage, H. M. (2018). A Qualitative Study of the Perceptions of Early Elementary Special Education Teachers on Flexible Seating and the Behavioral Needs of their Special Education Students. Dissertation. San Diego (CA): School of Northcentral University.

Schwarzer, R. (1992). *Self–Efficacy. Thought Control of Action*. Washington: Hemisphere.

Schwarzer, R., and Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: mediation analyses. *Appl. Psychol.* 57, 152–171. doi: 10.1111/j.1464-0597.2008.00359.x

Shankland, R., Benny, M., and Bressoud, N. (2017). Promotion de la santé mentale: les apports de la recherche en psychologie positive. *La Santé en Action.* 439, 17–19.

Sorrell, M. -A. (2019). Perceptions of flexible seating. *J. Teach. Act. Res.* 5, 120–136.

Tannenbaum, C., Ellis, R. P., Eyssel, F., Zou, J., and Schiebinger, L. (2019). Sex and gender analysis improves science and engineering. *Nature* 575, 137–146. doi: 10.1038/s41586-019-1657-6

Tiennot, S. (2019). Attention et assises flexibles. Master's Thesis. Rouen (FR): Université de Rouen.

Vallée, A. (2019). L'aménagement des classes du Québec: qu'en est-il? Master's Thesis. Trois-Rivières (QC): Université du Québec à Trois-Rivières.

Walker, S., and Berthelsen, D. (2007). Exploring gender differences in literacy and mathematical understandings in the early years of school: can differences be explained by behaviours in the classroom? Conference presentation. *Longitudinal Study of Australian Children Research Conference 2007, December*, Melbourne, Australia.

Wannarka, R., and Ruhl, K. (2008). Seating arrangements that promote positive academic and behavioural outcomes: a review of empirical research. *Support Learn.* 23, 89–93. doi: 10.1111/j.1467-9604.2008.00375.x

Welsh, J., Ford, L., Strazdins, L., and Friel, S. (2015). *Evidence Review: Addressing the Social Determinants of Inequities in Mental Wellbeing of Children and Adolescents*. Australia: Victorian Health Promotion Foundation.

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R., and Davis-Kean, P. (2006). "Motivation," in *Handbook of child psychology. 6th Edn*. ed. N. Eisenberg (United States: Wiley), 933–1002.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 06 July 2022
doi: 10.3389/fpsyg.2022.879178

STEM-Gender Stereotypes: Associations With School Empowerment and School Engagement Among Italian and Nigerian Adolescents

Pasquale Musso[1], *Maria Beatrice Ligorio*[1], *Ebere Ibe*[2]*, *Susanna Annese*[1], *Cristina Semeraro*[1] *and Rosalinda Cassibba*[1]

[1] Department of Educational Sciences, Psychology, Communication, University of Studies of Bari Aldo Moro, Bari, Italy,
[2] Department of Science Education, University of Nigeria, Nsukka, Nigeria

While many sociocultural, contextual, biological, behavioral, and psychological variables may contribute to the widespread under-representation of girls and women in the science, technology, engineering, and mathematics (STEM) field, this study focused on STEM-gender stereotypes, school experiences, and adolescence as critical factors in driving students' interest and motivation in STEM. Based on this, the study (a) investigated differences by gender and national context (Italy vs. Nigeria) in adolescents' STEM-gender stereotypes, school empowerment, and school engagement in a preliminary step, and (b) simultaneously examined how adolescents' STEM-gender stereotypes were related to school empowerment and school engagement as well as to socioeconomic status (SES). These latter relations were considered within the context of the potential moderating role of gender and national context. Participants included 213 Italian adolescents (M_{age} = 13.91; 52.1% girls) and 214 Nigerian adolescents (M_{age} = 13.92; 60.3% girls), who completed measures of school empowerment and engagement, STEM-gender stereotypes, and SES. A multivariate analysis of covariance showed that Nigerian girls and boys reported significantly higher levels of school empowerment, school engagement, and STEM-gender stereotypes than their Italian peers. Moreover, regardless of the national context, boys scored significantly higher on school empowerment and STEM-gender stereotypes than girls. Furthermore, a multiple-group path analysis revealed how higher school empowerment was related to lower STEM-gender stereotypes in both Italian and Nigerian girls' groups, while higher school engagement was associated with lower STEM-gender stereotypes only in the Nigerian groups. Regardless of gender and nationality, higher SES was linked to lower STEM-gender stereotypes. These findings particularly suggest that school empowerment and school engagement can be relevant dimensions to be studied and to develop strategies to counteract STEM-gender stereotypes in adolescence. Nonetheless, gender and national context are key factors to be considered. Limitations, strengths, future research, and educational implications are discussed.

Keywords: STEM-gender stereotypes, school empowerment, school engagement, socio-economic status, cultural comparison

OPEN ACCESS

Edited by:
Claudio Longobardi,
University of Turin, Italy

Reviewed by:
Giulia Savarese,
University of Salerno, Italy
Angelica Moè,
University of Padua, Italy

***Correspondence:**
Ebere Ibe
ebere.ibe@unn.edu.ng

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *18 February 2022*
Accepted: *30 May 2022*
Published: *06 July 2022*

Citation:
Musso P, Ligorio MB, Ibe E, Annese S, Semeraro C and Cassibba R (2022) STEM-Gender Stereotypes: Associations With School Empowerment and School Engagement Among Italian and Nigerian Adolescents. Front. Psychol. 13:879178. doi: 10.3389/fpsyg.2022.879178

INTRODUCTION

Science, technology, engineering, and mathematics (STEM) education is one of the key factors for preparing students for in-demand careers worldwide (e.g., Marginson et al., 2013; Zilberman and Ice, 2021; Eurostat, 2022). The continually evolving STEM sectors produce increasing opportunities to find entry-level work positions. This trend is proven not only in low- and middle-income countries, like those in Africa, where STEM education is strongly supported as a critical investment for social and economic development (World Bank, 2014) but also in high-income industrial countries. For example, in Italy 80% of STEM graduates find work within 1 year after graduation and this percentage becomes 92.1% within 5 years after graduation. These employment rates are significantly higher than those observed for graduates as a whole (AlmaLaurea, 2022) and confirm that new jobs are emerging within our economies, which require knowledge and skills in STEM. In addition to youth employment issues, global problems such as climate change, nutrition of a growing population, or growth of the economy itself can be better afforded by a new generation of well-educated young people in STEM. Therefore, STEM education has become a priority issue for both researchers from different fields and policymakers and non-governmental organizations.

However, despite this push toward the multiplication of actions to favor the spread of STEM education and employment, many countries are facing increasing gaps in this field (Kramer et al., 2015). As the United Nations Children's Fund (UNICEF) recently pointed out (2020a) in a specific report—much of this gap depends also on the under-representation of girls and women in STEM. The UNICEF report maps gender equity in STEM in 86 countries in different areas of the world and shows remarkable results. First, in more than 60% of the countries, girls at school (both at the upper primary and secondary level) show "minimum levels of proficiency" (MLP) in math and science at least comparable or higher than boys, but substantial differences exist depending on the regional area and socioeconomic status (SES). For example, girls present significantly lower MLP in math in most developing countries in Sub-Saharan Africa and Latin America, while within countries, girls show lower MLP in math than boys in the context of lower (but not higher) SES. Second, considering the "high proficiency levels" (HPL) in STEM, girls are less likely than boys to achieve these levels in most of the countries: 72% for math and 56% for science in upper primary school and 96% and 83%, respectively, at secondary school. Third, in around 60% of the countries, girls have a significant lower level of self-confidence in their STEM abilities than boys starting from the upper primary school; in the other 40% of countries, self-confidence scores also tend to be lower for girls, although in a statistically non-significant way. Fourth, girls' lower self-confidence is linked to a gender gap in STEM engagement, interest, and enjoyment, with correlations ranging from 0.44 to 0.65. Fifth, in 92% of countries, more boys than girls aspire to a STEM career; this gender gap is also evident even in the groups with the highest levels of STEM proficiency, with more than a fifth of boys aspiring to a STEM career in 64% of the countries, while this percentage drops to 17% for girls.

What emerges from this set of findings is that the under-representation of girls and women in the STEM field is generally widespread. This situation has evident negative consequences not only in terms of the development of the individual potential of half of the world's population (i.e., creativity, innovation, problem-solving, or increasing work-related STEM skills), but also from a more general, social, and political point of view. Without equal access and participation in STEM, for example, the 2030 Agenda for Sustainable Development provided by the United Nations (2015) will hardly reach its goals. STEM for girls, in fact, can stimulate and accelerate a number of Sustainable Development Goals (SDGs), like gender equality (SDG 5), no poverty (SDG 1), good health and wellbeing (SDG 3), and decent work and economic growth (SDG 8). This is because through better knowledge and use of science and technology (for example, related to health or communications) girls and women can potentially improve their lives and work-related opportunities. This acceleration may also concern quality education (SDG 4) or industry, innovation, and infrastructure (SDG 9), because higher STEM abilities empower girls to contribute to developing transferable, technical, and vocational skills for entrepreneurship and to lead innovative solutions in industrial sectors (UNICEF, 2020b).

In view of such a context and the potential negative future scenarios that arise from it, it is extremely important to understand why girls are under-represented in STEM and what actions can be taken to reverse the trend. While many sociocultural, contextual, biological, behavioral, and psychological factors may contribute to limiting girls' engagement with STEM, Master and Meltzoff (2016) highlighted the critical contribution of gender stereotypes in driving young students' interest and motivation in STEM. The under-representation of girls and women in the STEM field is deeply rooted in gender social representations that suggest how girls are not appropriate, or at least less than boys, for STEM education and employment (Master et al., 2014; Piatek-Jimenez et al., 2018; Thébaud and Charles, 2018; UNICEF, 2020b). Data from the above-mentioned report from UNICEF (2020a) support this view by associating gender gap in STEM with a variety of gender norms, biases, and stereotypes (e.g., girls receive less STEM-related praise; parents expect their sons, rather than their daughters, to have a STEM career). Regarding gender stereotypes, in many of the countries included in the report, 70% of individuals considers STEM as adequate for males than for females (e.g., Nosek et al., 2009; Campos et al., 2014; Cheryan et al., 2015; Grunspan et al., 2016; UNESCO, 2017; Schleicher, 2019).

Focusing on gender stereotypes is consistent with the most recent evidence claiming that the most important explanations of gender differences are grounded in preferences and choices rather than in skills and performance (e.g., Riegle-Crumb et al., 2012; Dasgupta and Stout, 2014). This approach explains why there would be fewer or no reliable gender differences in primary school than at later school levels when girls and boys more actively express preferences and interests and have been exposed to gender stereotypes influence for a longer time. In their work on these issues, Master and Meltzoff (2020) provide at least two key

suggestions. First of all, they distinguish between two dimensions of STEM-gender stereotypes (see also Master and Meltzoff, 2016; Wynn and Correll, 2017): a "cultural fit" stereotype (i.e., the belief that "STEM = male" and "girls like STEM less than boys") and an "ability" stereotype (i.e., the belief that "girls have less ability than boys"). Girls and women may worry about not fitting into the image of a STEM person and not having the ability to succeed in STEM and this combination contributes to their STEM under-representation. This broadens the concept of stereotype threat (Steele, 1997) and the related research approach, usually focused on how ability stereotypes affect girls' and women's performance in STEM and suggests using appropriate measures to grasp simultaneously "cultural fit" and "ability" stereotypes. Also, they propose a comprehensive STEreotypes, Motivation, and Outcomes (STEMO) developmental model, in which social factors (e.g., stereotypes) are essential in explaining youth's interest and academic outcomes in STEM. Specifically, this model indicates that when individuals encounter stereotypes about social groups (e.g., STEM-gender cultural fit and ability stereotypes) and these stereotypes are relevant to their social identity (e.g., gender), this has an impact on their self-representations (i.e., identification, ability beliefs, and sense of belonging) in STEM and, consequently, compromise their interest and academic achievement (e.g., participation) in STEM.

The STEMO model is a promising avenue for future interventions, given the centrality of STEM-gender stereotypes and their potential malleability in the school settings. From this point of view, one of the possible interventions is to challenge stereotypes about who belongs to STEM (cultural fit stereotypes) and the possession of fixed abilities determined by gender (ability stereotypes). According to the STEMO model, such interventions would have the consequence of changing the way girls would see themselves, increasing aspects such as the sense of identification with the STEM domain ("I am a math person"), the self-efficacy ("I am able to be successful in science and technology"), and the sense of belonging ("I am part of the STEM group"). This would lead to more positive STEM outcomes. Thus, one of the central questions is to evaluate which contextual, individual, social, and cultural factors favor overcoming the "traditional" STEM-gender stereotypes. In this study, we addressed the issue by focusing on (a) school context and adolescence; (b) two individual factors related to school experience, namely school empowerment and school engagement, theoretically associated with STEM-gender stereotypes; (c) one social factor like SES, given its influence on STEM outcomes (see above); and (d) gender and cultural differences, by comparing girls and boys from a high-income industrial country, such as Italy, with girls and boys from a low-middle-income country, such as Nigeria.

We considered that the processes suggested by the STEMO model unfold with the experiences in school (Master and Meltzoff, 2020), which represents one of the primary socialization environments for children and youth in terms of STEM subjects and expectations. Teachers and school staff may hold STEM-gender stereotypes influencing their interactions with students (e.g., Gunderson et al., 2012) as well as students' STEM-gender stereotypes and self-concepts (e.g., del Río et al., 2019). At higher school levels, students organize their tertiary educational and career preferences and choices also based on these experiences. Therefore, school is a privileged context to be considered both in terms of understanding the mechanisms that boost or buffer the transmission of STEM-gender stereotypes and in terms of potential interventions. In addition, adolescence represents a crucial life phase to be considered in relation to STEM-gender stereotypes. In fact, the most recent literature has adequately supported that traditional STEM-gender stereotypes were more prevalent among adolescents compared to younger children (e.g., Passolunghi et al., 2014; Miller et al., 2018; Starr and Simpkins, 2021). This finding was explained through the peculiarities of adolescence, a period when individuals are engaged in identity formation and try to use more systematically the information deriving from social confrontation (Erikson, 1968). Stereotypes may contribute to the development of identity because adolescents have the cognitive abilities to relate stereotypes to themselves (e.g., Marcia, 1994; Patterson and Bigler, 2018). Hence, adolescents represent a crucial group to be studied within the STEM-gender stereotypes research context.

Given the importance of the school context, dimensions such as school empowerment (Tam et al., 2020; Ruiz-Cantisani et al., 2021) and school engagement (Almeda and Baker, 2020) can play a role in the formation of STEM-gender stereotypes and STEM gender gap. Previous research suggested how there are links between STEM-gender stereotypes and self-efficacy: girls or women with higher explicit or implicit gender stereotypes in a STEM domain (e.g., math or science) frequently show lower beliefs to succeed in such a domain (e.g., Deemer et al., 2014; Passolunghi et al., 2014; Ertl et al., 2017). According to Bandura (1982), self-efficacy is subject-specific, and it should be conceptualized separately in each STEM domain. However, Zimmerman and Warschausky (1998) highlighted how self-efficacy is only a component of psychological empowerment, which "is not simply the belief that an individual can overcome barriers to independence, but also includes the individual's capacity and willingness to make such an effort (p. 13)." Cattaneo and Chapman (2010) argued that empowerment focuses on personally meaningful goals and aims to enhance one's social influence to exert power in social interaction. Starting from these conceptualizations, we assumed empowerment as a process that helps people gain control over their own lives (Page and Czuba, 1999), including reduced effects of stereotypes held by a society or community. The school community plays a relevant role in providing opportunities to experience psychological empowerment. Thus, higher levels of psychological empowerment experienced in the school context (school empowerment) could be related to lower levels of STEM-gender stereotypes, which usually reduce power in social influence and life choices. Furthermore, considering the STEM gender gap and the contents of STEM-gender stereotypes, it is still possible to assume a (negative) relation between school empowerment and STEM-gender stereotypes in girls, but not in boys.

The research also showed how students' engagement in the school context (school engagement) may be associated with STEM aspirations (Cunningham et al., 2015) and with the type of school programs chosen by the students, with those in

STEM programs more highly engaged than those in traditional programs (Patel et al., 2013; Kogo-Masila, 2017). However, the literature examining the association of school engagement with STEM-related dimensions is limited. To the best of our knowledge, no studies have investigated the relation between school engagement and STEM-gender stereotypes. Despite this paucity, there are reasons for this link to be explored. School engagement may be conceptualized as active, goal-directed, constructive interactions with the physical, social, and cultural environments of school (Furrer and Skinner, 2002) and, at a more individual level, may be operationalized as energy (i.e., positive approach), dedication (i.e., positive cognitive attitude), and absorption (i.e., concentration abilities) directed to school activities (Salmela-Aro and Upadyaya, 2012). Students who feel engaged with school show higher motivation and academic achievement over time (Salmela-Aro and Upadyaya, 2012); for girls, this dynamic may trigger greater curiosity and interest in STEM subjects as well as STEM-gender stereotype reactance with increased effort and willingness to demonstrate that the stereotypes are biased. Furthermore, this potential process can be more easily detectable in national contexts where the school still represents a concrete means for social redemption and where therefore school engagement can have more relevant outcomes from this point of view (i.e., more in a low-middle-income country rather than in a high-income country).

Social factors are also involved in students' STEM-gender stereotypes. Students from higher SES backgrounds may have previously been given more opportunities to learn about STEM and to build their STEM skills. This is especially important for girls, who can maximize their potential for success in STEM and, consequently, construct less biased STEM-gender stereotypes (Master and Meltzoff, 2020). On the contrary, girls from lower-SES backgrounds have fewer learning opportunities in STEM and chances to experience STEM skills; therefore, they may be more easily adherent to the culturally transmitted STEM-gender stereotypes.

Both STEM-gender stereotypes and the individual and social factors just described as well as their relations may vary depending on the students' gender and the national context of reference. STEM-gender stereotypes are cultural representations expressed by a particular society in many ways, such as social interactions and language use (Markus and Kitayama, 2010; Master and Meltzoff, 2020). They transcend beliefs within an individual, but when stereotypes concern issues involving gender, they can favor a gender more than another. Traditional STEM-gender stereotypes favor the boys, who may tend to conform to them less critically and present higher levels of stereotypes than girls, especially during adolescence (see Starr and Simpkins, 2021). Also, in low- and middle-income economies with higher levels of gender gap, STEM-gender stereotypes may be more prevalent (UNICEF, 2020b; World Economic Forum, 2021) than in developed countries, where all genders grow up by believing they share the same opportunities. Furthermore, previous studies reported consistent gender differences in school engagement, with girls more engaged with school than boys (e.g., Wang and Eccles, 2012; Fernández-Zabala et al., 2016). As Wang and Eccles (2012) reported, this finding may reflect a greater girls' concern for school performance, maybe because of gender socialization processes and differential expectations of parents and teachers (see also Wilkinson and Marrett, 1985; Eccles, 2007). The research also highlighted gender differences for school empowerment, with females scoring higher than males, and justified such a finding with the relevance of the social dimension for girls compared to boys (Helgeson, 1994; Årdal et al., 2018); yet these differences are small and further studies on this topic are needed. As for the differences related to the national context, to the best of our knowledge, literature does not report how school engagement and empowerment may change depending on country income levels (low and middle vs. high). However, it is theoretically possible that when school represents a greater opportunity for social mobility (in low- and middle-income countries), school engagement may be higher. Also, in terms of the relations among STEM-gender stereotypes, school empowerment, and school engagement, the research seems to be specifically lacking. Nevertheless, starting from the related literature, we previously suggested that (a) higher school empowerment may be associated with lower STEM-gender stereotypes in girls and (b) higher school engagement may be more associated with lower STEM-gender stereotypes in low-middle-income countries than in high-income countries, especially for girls. Finally, regarding the link between SES and STEM-gender stereotypes, previous research suggested that higher SES is associated with lower levels of STEM-gender stereotypes, and this is particularly evident for girls (Master and Meltzoff, 2020).

Aims and Hypotheses

In light of previous arguments, this study addressed the following two aims: (a) to assess gender and cultural differences in adolescents' STEM-gender stereotypes, school empowerment, and school engagement as a preliminary step; and (b) to analyze the associations of adolescents' school empowerment, school engagement, and SES with STEM-gender stereotypes and how these relations may change depending on gender and cultural context. To achieve these goals, as previously mentioned, we referred to two specific national contexts such as Italy and Nigeria, which are interesting to compare due to their socioeconomic and cultural characteristics. Italy is a European westernized country and one of the world's most industrialized economy with high-income levels. The gross domestic product (GDP) in Italy was 1,890 billion US dollars in 2020, according to official data from the World Bank (2022a). However, despite an improvement in the global gender gap index during the last 15 years, Italy is in the 63rd place across the 156 countries covered by the 2021 Global Gender Gap Report (GGGP, World Economic Forum, 2021) and presents a ratio of 1:0.46 in terms of STEM attainment in favor of males. Nigeria is a low-middle-income country, located in the western Sub-Saharan Africa. The GDP in Nigeria was 432 billion US dollars in 2020 (World Bank, 2022b). Nigeria experienced a slight improvement in the global gender gap index during the last 15 years as well, but it ranks 139th among the 156 countries (World Economic Forum, 2021). Although the 2021 GGGP does not report any indications about the male-female ratio of STEM attainment, a number

of reports have highlighted how Sub-Saharan Africa has one of the largest gender gaps worldwide in STEM, especially in the lower secondary school (e.g., Rubiano-Matulevich et al., 2019), and Nigeria presents a very low participation of females in STEM courses as a result of cultural and religious beliefs, traditions, early marriage, and parental educational background (e.g., Salman et al., 2011; Abdullahi et al., 2019).

Based on all the above information, we predicted that:

a) STEM-gender stereotypes were higher for boys than girls and for the Nigerian than the Italian adolescents.
b) School engagement was higher for girls than boys and in the Nigerian than in the Italian adolescents.
c) Higher school empowerment was significantly associated with lower STEM-gender stereotypes for girls, but not for boys.
d) Higher school engagement was more significantly associated with lower STEM-gender stereotypes in the Nigerian than in the Italian adolescents, especially for girls.
e) Higher SES was associated with lower levels of STEM-gender stereotypes, more significantly for girls than boys.

Given the lacking or less consistent literature as well as the exploratory nature of the study, we did not predict any specific gender and cultural differences for mean levels of school empowerment.

METHOD

Participants

The participants in this study included 213 Italian adolescents (M_{age} = 13.91, SD = 0.38, range = from 12 to 15 years; 47.9% boys and 52.1% girls) and 214 Nigerian adolescents (M_{age} = 13.92, SD = 0.97, range = from 12 to 15 years; 39.7% boys and 60.3% girls). Both the Italian and Nigerian participants attended the last year of lower secondary school; therefore, the following year, they would choose the higher education path that would lead them to a more restricted career perspective. This, therefore, represented a pivotal phase in their social identity development and sensitivity to STEM-gender stereotypes (see Introduction section). The Italian adolescents were attending school in southerneastern Italy (Apulia region) and Nigerian adolescents in southeastern Nigeria (Enugu State) in towns with more than 100,000 inhabitants. The average number of students in the classes frequented by the participants was 21.67 (SD = 3.59) for the Italian group and 38.19 (SD = 4.68) for the Nigerian group. The SES of the participants' families was prevalently medium. Based on a three-level classification of scores using the Barratt Simplified Measure of Social Status (BSMSS, Barratt, 2012, see Measures section), 4.2% of Italian and 9.3% of Nigerian adolescents fell into the low stratum, 62.0% of Italian and 58.9% of Nigerian adolescents fell into the medium stratum, and 33.8% of Italian and 31.8% of Nigerian adolescents fell into the high stratum. A comparison of the two national groups showed that they did not differ significantly in terms of gender (0 = boys, 1 = girls), $\chi^2(1) = 2.89$, $p = 0.09$, SES (0 = low, 1 = medium, 2 = high), $\chi^2(2) = 4.42$, $p = 0.11$, and age, $t(425) = -0.17$, $p = 0.87$. Significant differences were found for the average number of students in the classes, $t(425) = -40.98$, $p < 0.001$, with the Nigerian school classes more numerous than the Italian ones.

Procedure

The study was approved on 11 May 2020, by the Ethical Committee at the Department of Education, Psychology, and Communication at the University of Bari (Ethics reference code: ET-20-06), and all procedures were performed following the ethical principles for psychological research of the Italian Association of Psychology (2015). A convenience sample was initially recruited from three schools in the Italian urban context in Italy. The schools were selected by internal University search databases, where a list of local school institutions was stored, and encouraged to take part in the investigation through a motivation letter introducing the purpose of the research work. Within 1 month, the same procedure was followed in Nigeria by the third author of this work, who also ensured the comparability of the Nigerian schools with the Italian ones through a specific pairing process, by considering the schools' regional location in Nigeria and the urban characteristics in which they were inserted. After receiving permission from the respective school principals, the students' parents from both Italy and Nigeria were informed through a letter describing the purposes of the research, the voluntary nature of participation, and the anonymity of responses. All the parents provided informed consent for their son's or daughter's participation. In addition, participants provided signed assent agreeing to take part in the study. Participants completed a web-based survey in Italy and a web-based or a paper-and-pencil survey in Nigeria (depending on the schools) during the class time and they could withdraw at any time. The data collection took place between April and June 2021. Usually, participants completed the survey in about 30 min.

Measures

The measures used in this study were presented in the Italian language for the Italian participants, and in English for the Nigerian participants as it is the official and widely used language in Nigeria. When it was the case, we translated some measures from English into Italian (i.e., school engagement inventory). In the latter case, following the recommendations of the International Test Commission (2017), an independent English native language teacher, fluent in Italian, did a back-translation. Slight discrepancies were resolved through discussion and consensual agreement.

Socio-Demographics

Respondents were asked to indicate their age and gender. Paternal and/or maternal level of school completed (scores from 3 = *less than 7th grade* to 21 = *graduate degree*) and the parents' occupation (scores from 5 = *e.g., day laborer, house cleaner, food preparation worker* to 45 = *e.g., physician, judge, senior manager*) were assessed using BSMSS (Barratt, 2012; total education + total occupation scores from 8 to 66).

School Empowerment

An adapted form of the Psychological Empowerment Scale (PES; Spreitzer, 1995; see Pietrantoni and Prati, 2008, for

the Italian version) was used to assess students' perception of school empowerment. The original version of the PES consists of 12 items assessing four different dimensions in the workplace comprising three items each: meaning, competence, self-determination, and impact. However, recently it was adapted among students in different national contexts (e.g., Beauvais et al., 2014; Azizi et al., 2020; Cayaban et al., 2022). Following this line, we culturally adapted the instrument to students and the school environment. In doing so, the first three authors worked together following a specific procedure (see, for example, da Silva Augusto et al., 2017). Preliminarily, they discussed conceptual and semantic characteristics of PES, as previously adapted in the academic context, in light of the idiomatic and cultural differences (or equivalences) between the English and Italian versions as well as the Nigerian and Italian contexts. They agreed on the need to assess cross-culturally the PES content validity, which is the degree to which each item was relevant to and representative of school empowerment. Thus, they recruited a committee of six experts (three Italian and three Nigerian) with extensive experience in the school context, who rated each item on a Likert-type scale from 1 (*not important*) to 4 (*very important*). Only the items that obtained the maximum score (i.e., 4) from at least two Italian and two Nigerian experts were considered valid. Five items met this criterion, with at least one item in one of the four initial dimensions of the PES. After excluding (to maintain the item-dimension balance) the item with less agreement among experts, the final scale had four items: "The study I do is very important to me" for meaning, "I am confident about my ability to study" for competence, "I have opportunity for independence and freedom in how I study" for self-determination, and "I have significant influence over what happens in my class" for impact. Items were scored by the participants on a Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Prior studies have provided evidence that PES items load on four factors corresponding to the theoretical dimensions and that these factors load on a second-order factor of empowerment (e.g., Spreitzer, 1995; Pietrantoni and Prati, 2008). Thus, we expected that our four selected items would load on one factor of school empowerment across the two national contexts. We tested this one-factor structure model, as well as measurement invariance (configural, metric, and scalar, see Van de Schoot et al., 2012) across contexts, through robust maximum likelihood multi-group confirmatory factor analysis (MG-CFA; see the "Analysis Plan" section for model fit criteria), using the four items as observed indicators. This one-factor and scalar measurement-invariant model was adequately supported, $\chi^2(10) = 16.33$, $p = 0.09$, CFI = 0.970, RMSEA = 0.054, SRMR = 0.098. The internal consistency reliability scores calculated by the factor determinacy (Muthén and Muthén, 2012) were good for both the Italian (0.84) and Nigerian (0.78) groups. Overall, these results allowed validly comparing scale mean scores across the two national contexts (e.g., van de Vijver and Leung, 1997; Boer et al., 2018). For both groups, a composite variable was created by computing the average of the items, with higher scores indicating higher levels of school empowerment.

School Engagement

The Schoolwork Engagement Inventory (SEI; Salmela-Aro and Upadyaya, 2012) was used to assess students' perception of school engagement. The SEI consists of nine items assessing three different dimensions comprising three items each: energy, dedication, and absorption. To culturally adapt the instrument, we followed a procedure very similar to that already described for the school empowerment. The final scale had three items: "I feel strong and vigorous when I am studying" for energy, "I am enthusiastic about my studies." for dedication, and "Time flies when I am studying" for absorption. The items were scored by the participants on a Likert-type scale ranging from 1 (a *couple of times a year*) to 5 (*daily*). Prior studies have provided evidence that SEI items load better on one factor among the younger students (e.g., Salmela-Aro and Upadyaya, 2012). Following this line, we expected that our three selected items would load on one factor of school engagement across the two national contexts. We tested this one-factor structure model, as well as measurement invariance across contexts, through MG-CFA, using the three items as observed indicators. This one-factor and scalar measurement-invariant model was adequately supported, $\chi^2(4) = 7.75$, $p = 0.10$, CFI = 0.975, RMSEA = 0.066, SRMR = 0.085. The factor determinacy scores were good for both the Italian (0.94) and Nigerian (0.78) groups. For both groups, a composite variable was created by computing the average of the items, with higher scores indicating higher levels of school engagement.

STEM-Gender Stereotypes

To assess STEM-gender stereotypes, we used an eight-item questionnaire adapted by Tomasetto et al. (2015). This questionnaire measures explicit stereotypes concerning both between-gender (i.e., "I believe that generally males are more talented than females at math/science-technology") and within-gender (i.e., "I believe that generally females have more facility with language than with math/science-technology") differences in math (four items) and science-technology (four items). As it is possible to understand from the example items, the between-gender stereotypes recall the "ability" stereotypes, while the within-gender stereotypes recall the "cultural fit" stereotypes proposed by the STEMO model (Master and Meltzoff, 2020). To culturally adapt the instrument, we followed a procedure similar to that already described for the previous measures. All the items were retained. They were scored on a Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Prior studies have provided evidence that math-gender stereotypes items load on one factor (e.g., Tomasetto et al., 2015). Following this line, we expected that our eight items would load on two factor of math-gender and science/technology-gender stereotypes across the two national contexts. We tested this two-factor structure model, as well as measurement invariance across contexts, through MG-CFA, using the eight items as observed indicators. This two-factor and scalar measurement-invariant model was sufficiently supported, $\chi^2(46) = 107.71$, $p < 0.001$, CFI = 0.946, RMSEA = 0.079, SRMR = 0.094. The factor determinacy scores were good for both the Italian (0.96 and 0.97, respectively, for math-gender and science/technology-gender stereotypes) and Nigerian (0.92

and 0.94, respectively, for math-gender and science/technology-gender stereotypes) groups. However, the correlation between the two factors was very high: 0.95 for the Italian group and 0.78 for the Nigerian group. Based also on subsequent key analyses suggesting no differences in the patterns of results when considering math-gender and science/technology-gender stereotypes separately or as a whole, we used a unique variable of STEM-gender stereotypes henceforth for parsimony. The Cronbach's alpha coefficients for this general variable were: 0.96 for the Italian group and 0.85 for the Nigerian group. For both groups, a composite variable was created by computing the average of the eight items, with higher scores indicating higher levels of STEM-gender stereotypes.

Analytic Plan

The data analysis proceeded in three main steps. First, descriptive statistics for the study variables were initially calculated using version 24 of the *Statistical Package for the Social Sciences* (SPSS). Specifically, mean scores, standard deviations, normality statistics, and bivariate correlations were computed.

Second, we evaluated differences by gender (0 = boys; 1 = girls) and national context (0 = Italy; 1 = Nigeria) in school empowerment, school engagement, and STEM-gender stereotypes. Particularly, we conducted a multivariate analysis of covariance (MANCOVA) considering gender and national context as independent variables and the other constructs as dependent variables. SES was entered as a covariate.

Third, to explore the differential associations of school empowerment, school engagement, and SES with STEM-gender stereotypes and how these relations varied by gender and national context, a multiple-group path analysis using *Mplus 7* (Muthén and Muthén, 2012) was performed considering four groups: Italian boys, Italian girls, Nigerian boys, and Nigerian girls. We initially estimated and compared an unconstrained (less restrictive) model, in which the most relevant path coefficients were allowed to vary between the four groups, with a constrained (more restrictive) model, where all key path coefficients were set equal across groups. Significant differences in fit between these models implied the estimation of alternative partially constrained models. We relied on well-known goodness-of-fit indices and their associated cutoffs to evaluate the model fit (e.g., Kline, 2015): chi-square (χ^2) test with $p > 0.05$, CFI ≥ 0.90, RMSEA ≤ 0.08, and SRMR ≤ 0.10. To ascertain significant differences between nested models (the more vs. less restrictive model), at least two of these four criteria had to be satisfied (Kline, 2015): $\Delta\chi^2$ significant at $p < 0.05$, ΔCFI ≤ -0.010, ΔRMSEA ≥ 0.015, and ΔSRMR ≥ 0.010.

RESULTS

Preliminary Analyses

An initial data screening revealed that three participants (two Italians and one Nigerian) did not complete the survey (more than 30% of responses were not completed). These cases were deleted from the dataset. **Tables 1–3** summarize the descriptive statistics and report bivariate correlations in the total sample and by gender, by national context, and by gender and national context. They show how some observed variables were only slightly not normally distributed with skewness and kurtosis values $> \pm 1.00$ (Kline, 2015). This permitted us to perform the MANCOVA with some confidence, while in the structural equation modeling environment, the data were however analyzed using robust maximum likelihood estimation methods.

Mancova

Results from the MANCOVA showed a significant multivariate effect of gender, Wilks' Lambda = 0.95, $F_{(3,420)} = 7.94$, $p < 0.001$, $\eta^2 = 0.05$, and national context, Wilks' Lambda = 0.57, $F_{(3,420)} = 106.99$, $p < 0.001$, $\eta^2 = 0.43$. Two-way effects were not statistically significant. Follow-up univariate analyses (see **Table 4**) indicated that school empowerment and STEM-gender stereotypes differed significantly across gender, as well as school empowerment, school engagement, and STEM-gender stereotypes differed significantly across national contexts. Specifically, pairwise comparisons revealed that Nigerian participants reported significantly higher levels of all dependent variables than their Italian peers. Moreover, boys scored significantly higher on school empowerment and STEM-gender stereotypes than their female peers.

Multiple-Group Path Analysis

The theoretical model to be estimated across gender and national context is illustrated in **Figure 1**. The initial unconstrained model was a saturated model, $\chi^2(0) = 0.00$, $p = 0.00$, CFI = 1.00, RMSEA = 0.000, SRMR = 0.000. The constrained version of the model had poor fit, $\chi^2(18) = 74.62$, $p < 0.001$, CFI = 0.000, RMSEA = 0.172, SRMR = 0.175 and a significantly worse fit compared to the unconstrained model, $\Delta\chi^2(18) = 74.62$, $p < 0.001$, ΔCFI = −1.00, ΔRMSEA = 0.172, ΔSRMR = 0.175. Inspection of modification indices suggested releasing the constraints for (a) paths from school empowerment to STEM-gender stereotypes in the Italian male and Nigerian male groups, (b) paths from school engagement to STEM-gender stereotypes in the Italian groups compared to the Nigerian groups, and (c) covariances between school empowerment and school engagement in the Italian groups compared to the Nigerian groups. The obtained partially constrained model had excellent fit, $\chi^2(14) = 6.84$, $p = 0.94$, CFI = 1.00, RMSEA = 0.000, SRMR = 0.036 and did not have a significantly different fit compared to unconstrained model, $\Delta\chi^2(14) = 6.84$, $p = 0.94$, ΔCFI = 0.000, ΔRMSEA = 0.000, ΔSRMR = 0.036. Standardized coefficients of this final model are shown in **Figure 2**.

School empowerment was significantly and negatively related to STEM-gender stereotypes in both Italian and Nigerian female groups, while this association was significantly positive in the Italian male group and no significant relation was evidenced for the Nigerian male group. School engagement was significantly and negatively associated with STEM-gender stereotypes only in the Nigerian groups, while no significant relations were present in the Italian groups. SES was significantly and negatively linked to STEM-gender stereotypes in all considered groups. Furthermore, SES and school empowerment were significantly and positively correlated in all groups, while school empowerment and school

TABLE 1 | Means, standard deviations, skewness, and kurtosis for the key study variables for the entire sample, by gender, by national context and by gender and national context.

		M	*SD*	Skewness	Kurtosis
Entire sample ($N = 427$)					
1.	School empowerment (scored 1–5)	3.85	0.71	−0.34	−0.08
2.	School engagement (scored 1–5)	3.84	1.25	−0.96	−0.22
3.	STEM-gender stereotypes (scored 1–5)	2.83	1.19	0.00	−0.83
4.	Socio-economic status (scored 8–66)	48.42	14.29	−0.38	−0.92
Male group ($n = 187$)					
1.	School empowerment (scored 1–5)	3.93	0.71	−0.47	0.33
2.	School engagement (scored 1–5)	3.80	1.25	−0.92	−0.32
3.	STEM-gender stereotypes (scored 1–5)	3.00	1.27	−0.12	−0.95
4.	Socio-economic status (scored 8–66)	47.55	15.58	−0.41	−0.93
Female group ($n = 240$)					
1.	School empowerment (scored 1–5)	3.79	0.71	−0.26	−0.30
2.	School engagement (scored 1–5)	3.86	1.25	−0.99	−0.12
3.	STEM-gender stereotypes (scored 1–5)	2.70	1.12	0.04	−0.69
4.	Socio-economic status (scored 8–66)	49.11	13.20	−0.27	−1.13
Italian group ($n = 213$)					
1.	School empowerment (scored 1–5)	3.63	0.72	−0.37	0.03
2.	School engagement (scored 1–5)	3.31	1.32	−0.43	−1.09
3.	STEM-gender stereotypes (scored 1–5)	2.16	1.02	0.40	−0.62
4.	Socio-economic status (scored 8–66)	48.74	14.09	−0.41	−0.90
Nigerian group ($n = 214$)					
1.	School empowerment (scored 1–5)	4.07	0.64	−0.18	−0.78
2.	School engagement (scored 1–5)	4.36	0.91	−1.74	2.87
3.	STEM-gender stereotypes (scored 1–5)	3.50	0.95	−0.16	−0.48
4.	Socio-economic status (scored 8–66)	48.11	14.52	−0.35	−0.93
Italian male group ($n = 102$)					
1.	School empowerment (scored 1–5)	3.73	0.70	−0.54	0.75
2.	School engagement (scored 1–5)	3.40	1.34	−0.50	−1.11
3.	STEM-gender stereotypes (scored 1–5)	2.36	1.08	0.22	−0.63
4.	Socio-economic status (scored 8–66)	48.94	15.16	−0.61	−0.59
Italian female group ($n = 111$)					
1.	School empowerment (scored 1–5)	3.54	0.72	−0.23	−0.37
2.	School engagement (scored 1–5)	3.23	1.29	−0.39	−1.04
3.	STEM-gender stereotypes (scored 1–5)	1.97	0.93	0.48	−0.74
4.	Socio-economic status (scored 8–66)	48.55	13.10	−0.13	−1.46
Nigerian male group ($n = 85$)					
1.	School empowerment (scored 1–5)	4.17	0.64	−0.35	−0.64
2.	School engagement (scored 1–5)	4.29	0.91	−1.48	1.87
3.	STEM-gender stereotypes (scored 1–5)	3.75	1.04	−0.74	0.05
4.	Socio-economic status (scored 8–66)	45.87	16.00	−0.18	−1.15
Nigerian female group ($n = 129$)					
1.	School empowerment (scored 1–5)	4.00	0.63	−0.08	−0.78
2.	School engagement (scored 1–5)	4.40	0.91	−1.94	3.79
3.	STEM-gender stereotypes (scored 1–5)	3.34	0.85	0.21	−0.37
4.	Socio-economic status (scored 8–66)	49.59	13.31	−0.40	−0.83

engagement were significantly and positively associated only in the Italian groups and no significant relations were found in the Nigerian groups. No significant associations were revealed between SES and school engagement in all groups.

DISCUSSION

The purpose of the study was 2-fold. First, it investigated differences by gender and national context (Italy vs. Nigeria)

TABLE 2 | Pearson's bivariate correlations for the Italian sample.

		1.	2.	3.	4.
1.	School empowerment (scored 1–5)		0.48***	−0.14	0.18
2.	School engagement (scored 1–5)	0.51***		−0.02	0.01
3.	STEM-gender stereotypes (scored 1–5)	0.27**	0.12		−0.09
4.	Socio-economic status (scored 8–66)	0.18	0.09	−0.19	

*Upper diagonal: correlation matrix for females (n = 111). Lower diagonal: correlation matrix for males (n = 102). **p < 0.01, ***p < 0.001.*

TABLE 3 | Pearson's bivariate correlations for the Nigerian sample.

		1.	2.	3.	4.
1.	School empowerment (scored 1–5)		−0.01	−0.18*	0.21*
2.	School engagement (scored 1–5)	0.12		−0.24**	−0.09
3.	STEM-gender stereotypes (scored 1–5)	0.00	−0.20		−0.14
4.	Socio-economic status (scored 8–66)	0.26*	0.05	−0.03	

*Upper diagonal: correlation matrix for females (n = 129). Lower diagonal: correlation matrix for males (n = 85). *p < 0.05, **p < 0.01.*

TABLE 4 | Univariate analyses of covariance and pairwise comparisons for gender and national context (Italian vs. Nigerian) on school empowerment, school engagement, and STEM-gender stereotypes.

	MANCOVA-adjusted means by gender				MANCOVA-adjusted means by national context			
	Male (n = 187)	Female (n = 240)	F(1, 422)	η^2	Italian (n = 213)	Nigerian (n = 214)	F(1, 422)	η^2
School empowerment	3.96[a]	3.76[b]	9.21**	0.02	3.63[a]	4.09[b]	50.66***	0.11
School engagement	3.84	3.82	0.06	0.00	3.31[a]	4.35[b]	86.83***	0.17
STEM-gender stereotypes	3.05[a]	2.66[b]	17.30***	0.04	2.17[a]	3.54[b]	210.88***	0.33

*A mean is significantly different (p < 0.05) from another mean within the same row if they have different superscripts (a or b). **p < 0.01, ***p < 0.001. MANCOVA, multivariate analysis of covariance.*

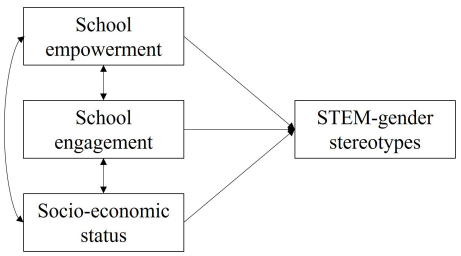


FIGURE 1 | The theoretical model to be estimated across gender and national context (Italian males, Italian females, Nigerian males, and Nigerian females).

in adolescents' STEM-gender stereotypes, school empowerment, and school engagement. Second, and more importantly, for the first time, it simultaneously analyzed how adolescents' STEM-gender stereotypes are related to the individual resources of school empowerment and school engagement as well as to the social factor of SES. These relations were considered in the context of the potential moderating role of gender and national context. The main results revealed that boys outscored girls in STEM-gender stereotypes and school empowerment and that Nigerian adolescents outperformed the Italian adolescents in STEM-gender stereotypes, school empowerment, and school engagement. Furthermore, higher school empowerment was significantly associated with lower STEM-gender stereotypes for girls regardless of the national context, while higher school engagement was associated with lower STEM-gender stereotypes in the Nigerian groups. Higher SES was associated with lower levels of STEM-gender stereotypes regardless of gender and national context. These results might suggest that, in addition to SES, school empowerment and school engagement can be relevant to be studied and to develop strategies to counteract STEM-gender stereotypes in adolescence. Nonetheless, it is necessary to consider the role of gender and national context to provide a better and appropriate interpretation of the emerging dynamics.

Gender and Cultural Differences in Adolescents' STEM-Gender Stereotypes, School Empowerment, and School Engagement

We expected higher levels of STEM-gender stereotypes for boys than girls and for the Nigerian than the Italian contexts. Our findings supported this prediction. As previous literature

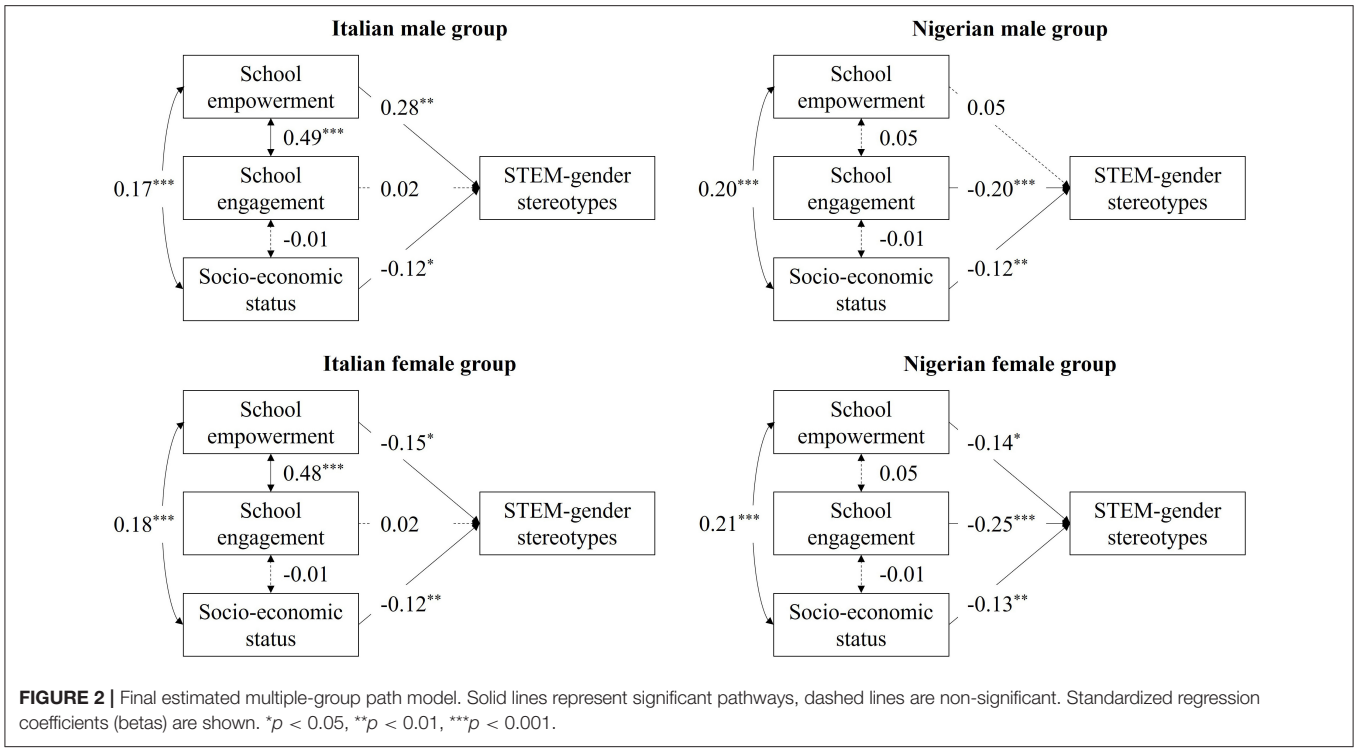


FIGURE 2 | Final estimated multiple-group path model. Solid lines represent significant pathways, dashed lines are non-significant. Standardized regression coefficients (betas) are shown. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

extensively reported (e.g., Moè et al., 2021; Starr and Simpkins, 2021), in adolescence, boys endorse STEM-gender stereotypes more strongly than girls. One explanation for this is that generally people conform more easily to associations that favor their gender. STEM-gender stereotypes propose associations that favor boys (e.g., "STEM = male"), while disadvantaging girls ("girls have less ability than boys"). This process of "favoritism" could, in turn, increase the perception of congruity between boys' gender role (what choices and behaviors they consider typical for their gender) and the beliefs that members of society usually have about what is most appropriate for them (Eagly and Karau, 2002), further fostering their stereotypes. Favoritism and gender role congruence, therefore, may account for the higher levels of STEM-gender stereotypes in boys than girls.

The national context, along with related cultural and social features, is also a principal factor that differentiates the levels of STEM-gender stereotypes. In line with prior research (UNICEF, 2020b; World Economic Forum, 2021), we found much higher levels of STEM-gender stereotypes in Nigeria than in Italy. Nigeria is a low-middle-income country with a high gender gap. It is still facing serious issues regarding gender differences (Salman et al., 2011; Abdullahi et al., 2019), linked to religious dimensions, cultural traditions (e.g., early marriages for girls), and socio-political issues (e.g., the general education levels of the population). In this context, STEM-gender stereotypes are widespread in the Nigerian population, resulting in a greater inequality of opportunities between boys and girls. Looking at the 2021 GGGP data (World Economic Forum, 2021), these sociocultural processes seem less relevant in Italy, where boys and girls can have a more equal view of future life and professional chances. This could explain why Italian boys and girls have lower levels of STEM-gender stereotypes.

We also expected higher levels of school engagement for girls than boys as well as for the Nigerian than the Italian adolescents. The findings supported this prediction only partially. First, as hypothesized, our adolescent participants in Nigeria showed higher levels of school engagement. When a context generally offers fewer prospects for personal life and profession, as in Nigeria than in Italy, the school may be perceived as one of the most significant and catalyzing environments providing opportunities for social climbing. This can lead to living in school in a more active and energetic way, building more positive attitudes toward the academic experience, fueling greater concentration in achieving goals, and ensuring more meaningful social relationships (Furrer and Skinner, 2002; Salmela-Aro and Upadyaya, 2012). Second, our expectation of higher levels of school engagement for girls than boys was not supported, and no differences were found. This is not consistent with previous research, which suggested that personality and motivation factors (e.g., Lam et al., 2012) or differential expectations of parents and teachers (Wilkinson and Marrett, 1985; Eccles, 2007; Wang and Eccles, 2012) may promote girls' greater concern on their school connection and performance. Probably, this result should be interpreted in relation to our group of participants and the period of data collection. As mentioned above, all our participants attended the last year of lower secondary school in the last period of the school year, when shortly thereafter they would face the final exams and the choice of the higher education path. This may have favored a general greater engagement by all students toward the final goal, flattening any inter-individual and gender

differences. Therefore, further studies with larger samples, at different school grades, and at various times of the school year would be desirable, especially if the design is longitudinal.

We took an exploratory approach in considering gender and cultural differences in mean levels of school empowerment. Boys and Nigerian participants showed higher school empowerment than girls and Italian participants. In terms of gender differences, a previous study showed significantly higher mean scores of girls than boys for school empowerment, but the effect size was small (Årdal et al., 2018). However, the school empowerment measure was not strictly comparable to that of our study. Årdal et al. (2018) used a measure referring to motivation for influencing school, perceived control, and participatory behavior (Ozer and Schotland, 2011). Our measure was related to meaning, competence, self-determination, and impact. The main difference can be identified in the inclusion, in our measure, of the competence dimension, for which boys usually score higher than girls (e.g., Conway et al., 2015; Gomez-Baya et al., 2019). This can at least partially explain our result. Nevertheless, this finding raises the question of whether social norms and cultural stereotypes can have a strong impact on girls, inhibiting those empowerment and assertiveness skills crucial for the promotion of their interests and demands (Hentschel et al., 2019). This topic should be addressed in future research. Regarding the higher levels of school empowerment of Nigerian students compared to Italian ones, this again seems to support the idea that in Nigeria, more than in Italy, school seems to be a significant and catalyzing context for the expression of the individual resources of boys and girls, who seem to be better able to experience school as a setting of active responsibility.

Associations of Adolescents' STEM-Gender Stereotypes With School Empowerment, School Engagement, and SES in the Context of the Moderating Role of Gender and National Context

Concerning our primary goal, the findings showed that our expectations were generally supported with some exceptions. As expected, higher school empowerment was associated with lower levels of STEM-gender stereotypes in the two groups of girls regardless of the national context. Higher levels of school empowerment contribute to giving girls more control over their lives (Page and Czuba, 1999), by focusing on personal goals and enhancing their power in social interaction (Cattaneo and Chapman, 2010). This can make it easier for girls to react to the socially widespread STEM-gender stereotypes, which in contrast reduce their active self-determination and participation. The dynamic characterizing the boys is different. No significant association between school empowerment and STEM-gender stereotypes was evidenced for the Nigerian boys, while the association was positive for the Italian boys. Given that boys belong to the "gender favored by STEM-gender stereotypes," it is not relevant for them to refer to empowering processes to counteract their social beliefs. On the contrary, the active management of social power might favor increasing levels of STEM-gender stereotypes. In line with this argument, it makes sense to expect this second mechanism to emerge in national contexts with a greater rate of individuality and where personal goals and success take on high relevance, such as in Italy, rather than in more collectivist contexts, such as Nigeria (Hofstede, 2001).

As far as the relations between school engagement and STEM-gender stereotypes are concerned, we found significantly negative links in the two groups of Nigerian boys and girls, while no associations were evidenced in the Italian groups. We hypothesized these differences related to the context, but we also expected some differences concerning gender. More specifically, we assumed a negative link between school engagement and STEM-gender stereotypes in the Italian girls, albeit less strong than that of the Nigerian girls. When students feel particularly engaged with the school, they are more inclined to consider it as a source of personal improvement, support, and motivation (Salmela-Aro and Upadyaya, 2012). This expands one's enthusiasm and interest also toward fields that social stereotypes would suggest as unsuitable and motivates to oppose these stereotypes. Such a process could therefore explain how school engagement would help reduce STEM-gender stereotypes in the Nigerian girls. However, this process might interact with the national context of reference. The more the school is considered a social value in terms of opportunities for social mobility in low-middle-income contexts, such as Nigeria, the more this process could be relevant and unfold its effects. In contexts where the levels of economic development and social support are higher, such as in Italy, the school could instead be perceived as a less determining factor for future subsistence, and this could dampen the fundamental meaning of the hypothesized process. This could be one reason for the lack of significant relation between school engagement and STEM-gender stereotypes in Italian girls. The two Italian and Nigerian contexts probably differ in another aspect as well. In the Nigerian context, where greater gender gaps and STEM-gender stereotypes are present, greater school engagement may imply greater attention to information counteracting these issues at school, and to girls when they present clear STEM skills. This could explain why school engagement was negatively associated for Nigerian boys in an equally relevant way as for Nigerian girls, while no significant relation was found for Italian boys, living in a context characterized by significantly lower levels of gender gap and STEM-gender stereotypes than in Nigeria (and this may make boys generally less sensitive to information and experiences promoting gender equity at school).

Finally, higher SES is related to lower levels of STEM-gender stereotypes, regardless of gender and national context. We expected this finding to be particularly relevant for girls compared to boys. In fact, we thought that higher levels of SES provided better chances for STEM learning and skills. Such a situation could more easily lead girls to reduce STEM-gender stereotypes than boys. However, our results suggest that SES background is equally relevant for boys as well. Although the literature suggests that STEM-gender stereotypes are more prevalent among adolescents than younger children due to advanced cognitive abilities connecting their identity with social categories (e.g., Passolunghi et al., 2014; Miller et al., 2018; Starr

and Simpkins, 2021), however, other changes related to critical and moral skills could be generally associated with lower levels of STEM-gender stereotypes (e.g., Malti et al., 2021). A higher SES background could foster such skills for both girls and boys, and this could more easily explain our findings showing a lack of gender differences.

Limitations, Strengths, and Future Research

This study should be considered in light of some weaknesses. First, we used a convenience sampling method to collect our research data, and this casts doubt on the generalizability of our results. Also, because of selection bias, it is possible that the schools that participated in the study were significantly more motivated and/or more satisfied with their education paths and activities than those which did not. Large population-based random samples would be ideal to be considered in future research. Second, the use of self-report measures requires caution when interpreting the findings, even more when diverse cultural contexts are considered. Next investigations should combine mixed methods. For example, the simultaneous use of qualitative and quantitative analysis could help highlight the subjective experience of boys and girls in various national contexts. Third, the cross-sectional nature of the study design precludes us from clearly concluding the direction of the associations among the study variables (for example, from school empowerment to STEM-gender stereotypes or vice versa). Thus, it would be important to conduct future longitudinal studies following the same participants during adolescence in order to draw clearer conclusions about the direction of associations between these variables and about the causality processes involved. Fourth, our study was limited to the investigation of the associations of school empowerment, school engagement, and SES with STEM-gender stereotypes within the context of potential differences by gender and nationality. Actually, other variables may be interesting to consider. For instance, further studies could consider how the family environment and parenting, peer experiences, teacher-student relationships, sense of community, and personal future expectations could directly or indirectly affect STEM-gender stereotypes (e.g., Tandrayen-Ragoobur and Gokulsing, 2021). Furthermore, it is noteworthy to point out that we focused on explicit stereotypes only, namely on conscious representations assessed through self-reports, which may produce biased responses due to social desirability. To prevent such concerns, many studies analyzed the role of implicit gender stereotypes in STEM performance (e.g., Hausmann, 2014). This suggests that future research should consider assessing both implicit and explicit stereotypes and comparing the results.

Despite these limitations, our study contributed meaningfully to the literature because it extends our understanding of the characteristic of STEM-gender stereotypes in two ways. First, it provided a new clear picture of how STEM-gender stereotypes may differ based on gender and nationality. Second, it revealed how significant and school-based variables (empowerment and engagement) are associated with STEM-gender stereotypes, considering the role of gender and national contexts in these relations. Together, the findings highlighted potential factors to work on to reduce STEM-gender stereotypes from an international perspective. However, interventions should be developed by taking into account gender and national differences.

Educational Implication

Our findings provide implications for practice in the school community. Based on the STEMO model, we considered STEM-gender stereotypes as composed by two dimensions, i.e., cultural fit and ability stereotypes. To reduce the impact of these two types of stereotypes, it is important, on the one hand, to think of interventions that broaden the idea of who can be part of the STEM field and, on the other, to counteract the idea that skills are fixed (Master and Meltzoff, 2020). In the first case, for example, one could work by making the school environments dedicated to STEM teaching (for example, the computer room or the chemistry lab) less stereotypically masculine, using expedients such as the presence of plants, furniture with fluid lines, and colors usually matched to the feminine style (for example, powder pink and lilac). In the second case, it would be important for teachers to convey the idea that STEM skills are like a sporting activity: the more you practice and train, the more results will be obtained. Emphasizing the initial mistakes and failures of great scientists, who then achieved success by working hard, can be a good strategy. Moreover, in this line, Law et al. (2021) reported a good example of a growth mindset activity in a science museum. Both interventions seem particularly crucial to practice in contexts with high levels of stereotypes, such as Nigeria. Furthermore, they should be systematically addressed not only to girls, but also to boys, who are the holders of the highest levels of STEM-gender stereotypes and, therefore, as future fathers or managers, could hinder the STEM interests or careers of girls and women.

Based on our findings, it would also be important to design interventions to boost girls' school empowerment. To achieve this goal, schools and teachers should be committed to providing them with meaningful school environments, feelings of confidence in school work, opportunities for self-determination, and a sense of impact at school. Motivation training, aimed at making girls more confident and perceiving themselves as more able and capable to increase their performance, have proved to be effective interventions and deserve to be replicated (e.g., Moè, 2016). Another important strategy could be to provide positive role models in the use of empowerment skills (Master and Meltzoff, 2016). Teachers might represent such positive role models (they do not necessarily have to be females, just relatable and similar to the self along certain key dimensions), but also schoolmates who "are like me and manage to be influential and achieve their goals" can fulfill a similar function. Furthermore, simple activities such as assigning responsibility for leading teamwork could prove effective and easily applicable.

Intervention programs should also promote school engagement, being aware that such interventions are likely to be most effective where the value of the school is generally believed to be more socially crucial, such as in Nigeria compared to Italy. One way to achieve this goal is to strengthen the sense of belonging to the school, by proposing activities that reinforce the idea that school is a "meaningful context of life." In this line, the redefinition of academic programs toward topics close to students' experiences, the offer of extracurricular activities of interest to them (concerning, for example, sports and music), the support of significant peer tutors, and motivational programs could be important.

It should be noted that all the interventions previously outlined involve the school microsystem of girls and boys. However, our study suggests that other factors related to the socioeconomic and cultural development of home nations and families also potentially play a role in the formation of STEM-gender stereotypes. At this level, economic and social policy interventions are desirable in the direction of providing more girls and boys with opportunities for knowledge and experiences in STEM. Such occasions should suggest that the involvement of both genders in the STEM field is a crucial point for the wellbeing and progress of our living communities.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available upon reasonable request by the authors, without undue reservation.

ETHICS STATEMENT

The study was approved on 11 May 2020, by the Ethical Committee at the Department of Education, Psychology, and Communication at the University of Bari (Ethics reference code: ET-20-06). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

PM provided the conception and design of this study, performed and interpreted the data analyses, and wrote the first draft of the manuscript. ML and EI contributed to the conception of this study, collected data, drafted, and revised the article substantially. SA and CS revised the article and proposed important suggestions for modification. RC revised the article critically for important intellectual content. All authors contributed to the article and approved the submitted version.

REFERENCES

Årdal, E., Holsen, I., Diseth, Å., and Larsen, T. (2018). The five Cs of positive youth development in a school context; gender and mediator effects. *Sch. Psychol. Int.* 39, 3–21. doi: 10.1177/0143034317734416

Abdullahi, N., Abubakar, A., Abubakar, M. J., and Aliyu, A. C. (2019). Gender gap in science and technology education in Nigeria. *Int. J. Educ. Eval.* 5, 6–13. Available online at: https://www.iiardjournals.org/get/IJEE/VOL.%205%20NO.%203%202019/Gender%20Gap%20in%20Science.pdf

AlmaLaurea (2022). *Laureate e laureati: scelte, esperienze e realizzazioni professionali. Rapporto 2022* [Female and male graduates: Choices, experiences and professional achievements. 2022 Report]. Available online at: https://www.almalaurea.it/universita/altro/2022/rapporto-laureate-laureati (accessed April 10, 2022).

Almeda, M. V. Q., and Baker, R. S. (2020). Predicting student participation in STEM careers: The role of affect and engagement during middle school. *J. Educ. Data Mining* 12, 33–47. doi: 10.5281/zenodo.4008054

Azizi, S. M., Heidarzadi, E., Soroush, A., Janatolmakan, M., and Khatony, A. (2020). Investigation the correlation between psychological empowerment and assertiveness in nursing and midwifery students in Iran. *Nurse Educ. Pract.* 42, 102667. doi: 10.1016/j.nepr.2019.102667

Bandura, A. (1982). Self-efficacy mechanism in human agency. *Am. Psychol.* 37, 122–147. doi: 10.1037/0003-066X.37.2.122

Barratt, W. (2012). *The Barratt Simplified Measure of Social Status (BSMSS)*. Available online at: http://socialclassoncampus.blogspot.it/2012/06/barratt-simplified-measure-of-social.html

Beauvais, A. M., Stewart, J. G., DeNisco, S., and Beauvais, J. E. (2014). Factors related to academic success among nursing students: A descriptive correlational research study. *Nurse Educ. Today* 34, 918–923. doi: 10.1016/j.nedt.2013.12.005

Boer, D., Hanke, K., and He, J. (2018). On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests. *J. Cross Cult. Psychol.* 49, 713–734. doi: 10.1177/0022022117749042

Campos, F., Goldstein, M., McGorman, L., Boudet, A. M., and Pimhidzai, O. (2014). *Breaking the Metal Ceiling: Female Entrepreneurs Who Succeed in Male-Dominated Sectors in Uganda*. Africa Region Gender Practice Policy Brief, 9. Available online at: https://openknowledge.worldbank.org/bitstream/handle/10986/25456/938940BRI00PUB0ng0the0Metal0Ceiling.pdf?sequence=1&isAllowed=y (accessed April 12, 2022).

Cattaneo, L. B., and Chapman, A. R. (2010). The process of empowerment: A model for use in research and practice. *Am. Psychol.* 65, 646–659. doi: 10.1037/a0018854

Cayaban, A. R. R., Valdez, G. F. D., Leocadio, M. L., Cruz, J. P., Tuppal, C. P., Labrague, L. J., et al. (2022). Structural and psychological empowerment and its influencing factors among nursing students in Oman. *J. Professional Nursing* 39, 76–83. doi: 10.1016/j.profnurs.2022.01.003

Cheryan, S., Master, A., and Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes. *Front. Psychol.* 6, 49. doi: 10.3389/fpsyg.2015.00049

Conway, R. J., Heary, C., and Hogan, M. J. (2015). An evaluation of the measurement properties of the five Cs model of positive youth development. *Front. Psychol.* 6, 1941. doi: 10.3389/fpsyg.2015.01941

Cunningham, B. C., Hoyer, K. M., and Sparks, D. (2015). Gender *Differences in Science, Technology, Engineering, and Mathematics (STEM) Interest, Credits Earned, and NAEP Performance in the 12th Grade. Statistics in Brief*. NCES 2015-075. National Center for Education Statistics.

da Silva Augusto, F., Blanes, L., Nicodemo, D., and Ferreira, L. M. (2017). Translation and cross-cultural adaptation of the Cardiff Wound Impact Schedule to Brazilian Portuguese. *J. Tissue Viability* 26, 113–118. doi: 10.1016/j.jtv.2016.12.002

Dasgupta, N., and Stout, J. G. (2014). Girls and women in science, technology, engineering, and mathematics: STEMing the tide and broadening participation in STEM careers. *Policy Insights Behav. Brain Sci.* 1, 21–29. doi: 10.1177/2372732214549471

Deemer, E. D., Thoman, D. B., Chase, J. P., and Smith, J. L. (2014). Feeling the threat: Stereotype threat as a contextual barrier to women's science career choice intentions. *J. Career Dev.* 41, 141–158. doi: 10.1177/0894845313483003

del Río, M. F., Strasser, K., Cvencek, D., Susperreguy, M. I., and Meltzoff, A. N. (2019). Chilean kindergarten children's beliefs about mathematics: Family matters. *Dev. Psychol.* 55, 687–702. doi: 10.1037/dev0000658

Eagly, A. H., and Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* 109, 573–598. doi: 10.1037/0033-295X.109.3.573

Eccles, J. S. (2007). "Where are all the women? Gender differences in participation in physical science and engineering," in *Why Aren't More Women in Science?*

eds S. J. Ceci and W. M. Williams. Washington, DC: American Psychological Association. doi: 10.1037/11546-016

Erikson, E. H. (1968). *Identity: Youth and Crisis*. New York, NY: Norton.

Ertl, B., Luttenberger, S., and Paechter, M. (2017). The impact of gender stereotypes on the self-concept of female students in STEM subjects with an under-representation of females. *Front. Psychol.* 8, 703. doi: 10.3389/fpsyg.2017.00703

Eurostat (2022). *Human Resources in Science and Technology*. Available online at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Human_resources_in_science_and_technology#Professionals_and_technicians_employed_in_science_and_technology_occupations (accessed April 10, 2022).

Fernández-Zabala, A., Goñi, E., Camino, I., and Zulaika, L. M. (2016). Family and school context in school engagement. *Eur. J. Educ. Psychol.* 9, 47–55. doi: 10.1016/j.ejeps.2015.09.001

Furrer, C., and Skinner, E. (2002). Sense of relatedness as a factor in children's academic engagement and performance. *J. Educ. Psychol.* 95, 148–162. doi: 10.1037/0022-0663.95.1.148

Gomez-Baya, D., Reis, M., and Gaspar de Matos, M. (2019). Positive youth development, thriving and social engagement: An analysis of gender differences in Spanish youth. *Scand. J. Psychol.* 60, 559–568. doi: 10.1111/sjop.12577

Grunspan, D. Z., Eddy, S. L., Brownell, S. E., Wiggins, B. L., Crowe, A. J., and Goodreau, S. M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PLoS ONE* 11, e0148405. doi: 10.1371/journal.pone.0148405

Gunderson, E. A., Ramirez, G., Levine, S. C., and Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles* 66, 153–166. doi: 10.1007/s11199-011-9996-2

Hausmann, M. (2014). Arts versus science—Academic background implicitly activates gender stereotypes on cognitive abilities with threat raising men's (but lowering women's) performance. *Intelligence* 46, 235–245. doi: 10.1016/j.intell.2014.07.004

Helgeson, V. S. (1994). Prototypes and dimensions of masculinity and femininity. *Sex Roles* 31, 653–682. doi: 10.1007/BF01544286

Hentschel, T., Heilman, M. E., and Peus, C. V. (2019). The multiple dimensions of gender stereotypes: A current look at men's and women's characterizations of others and themselves. *Front. Psychol.* 10, 11. doi: 10.3389/fpsyg.2019.00011

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. Beverly Hills, CA: Sage

International Test Commission (2017). *The ITC Guidelines for Translating and Adapting Tests (Second edition)*. Available online at: https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf.

Italian Association of Psychology (2015). *Codice etico per la ricerca in psicologia* [Ethical code for psychological research]. Available online at: http://www.aipass.org/node/11560.

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford.

Kogo-Masila, C. A. (2017). *A Comparative Study of Two Graduation Pathways: Traditional vs. STEM at a Southeastern High School* [Unpublished doctoral dissertation]. Gardner-Webb University, Boiling Springs, NC. Available online at: https://digitalcommons.gardner-webb.edu/cgi/viewcontent.cgi?article=1246&context=education_etd

Kramer, M., Tallant, K., Goldberger, A., and Lund, F. (2015). *The Global STEM Paradox*. New York, NY: New York Academy of Sciences.

Lam, S. F., Jimerson, S., Kikas, E., Cefai, C., Veiga, F. H., Nelson, B., et al. (2012). Do girls and boys perceive themselves as equally engaged in school? The results of an international study from 12 countries. *J. School Psychol.* 50, 77–94. doi: 10.1016/j.jsp.2011.07.004

Law, F., McGuire, L., Winterbottom, M., and Rutland, A. (2021). Children's gender stereotypes in STEM following a one-shot growth mindset intervention in a science museum. *Front. Psychol.* 12, 1602. doi: 10.3389/fpsyg.2021.641695

Malti, T., Galarneau, E., and Peplak, J. (2021). Moral development in adolescence. *J. Res. Adolesc.* 31, 1097–1113. doi: 10.1111/jora.12639

Marcia, J. E. (1994). "The empirical study of ego identity," in *Identity and Development: An Interdisciplinary Approach*, eds H. A. Bosma, T. G. Graafsma, H. D. Grotevant, and D. J. de Levita. Thousand Oaks, CA: Sage.

Marginson, S., Tytler, R., Freeman, B., and Roberts, K. (2013). STEM: *Country Comparisons. International Comparisons of Science, Technology, Engineering and Mathematics (STEM) Education*. Final report. Australian Council of Learned Academies.

Markus, H. R., and Kitayama, S. (2010). Cultures and selves: A cycle of mutual constitution. *Perspect. Psychol. Sci.* 5, 420–430. doi: 10.1177/1745691610375557

Master, A., Cheryan, S., and Meltzoff, A. N. (2014). Reducing adolescent girls' concerns about STEM stereotypes: When do female teachers matter?. *Revue Int. Psychol. Soc.* 27, 79–102. Available online at: https://www.cairn.info/revue-internationale-de-psychologie-sociale-2014-3-page-79.htm

Master, A., and Meltzoff, A. N. (2016). Building bridges between psychological science and education: Cultural stereotypes, STEM, and equity. *Prospects* 46, 215–234. doi: 10.1007/s11125-017-9391-z

Master, A., and Meltzoff, A. N. (2020). Cultural stereotypes and sense of belonging contribute to gender gaps in STEM. *Int. J. Gender Sci. Technol.* 12, 152–198. Available online at: http://genderandset.open.ac.uk/index.php/genderandset/article/view/674

Miller, D. I., Nolla, K. M., Eagly, A. H., and Uttal, D. H. (2018). The development of children's gender science stereotypes: A meta-analysis of 5 decades of U.S. draw-a-scientist studies. *Child Dev.* 89, 1943–1955. doi: 10.1111/cdev.13039

Moè, A. (2016). Teaching motivation and strategies to improve mental rotation abilities. *Intelligence* 59, 16–23. doi: 10.1016/j.intell.2016.10.004

Moè, A., Hausmann, M., and Hirnstein, M. (2021). Gender stereotypes and incremental beliefs in STEM and non-STEM students in three countries: Relationships with performance in cognitive tasks. *Psychol. Res.* 85, 554–567. doi: 10.1007/s00426-019-01285-0

Muthén, L. K., and Muthén, B. (2012). *1998-2012. Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., et al. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proc. Nat. Acad. Sci. U.S.A.* 106, 10593–10597. doi: 10.1073/pnas.0809921106

Ozer, E. J., and Schotland, M. (2011). Psychological empowerment among urban youth: Measure development and relationship to psychosocial functioning. *Health Educ. Behav.* 38, 348–356. doi: 10.1177/1090198110373734

Page, N., and Czuba, C. E. (1999). Empowerment: What is it. *J. Extension* 37, 1–5.

Passolunghi, M. C., Ferreira, T. I. R., and Tomasetto, C. (2014). Math–gender stereotypes and math-related beliefs in childhood and early adolescence. *Learn. Individ. Differ.* 34, 70–76. doi: 10.1016/j.lindif.2014.05.005

Patel, N. H., Franco, M. S., and Lindsey, J. (2013). Examining student engagement in open enrollment STEM-focused schools. *Open J. Educ.* 1, 52–60. doi: 10.12966/oje.05.07.2013

Patterson, M. M., and Bigler, R. S. (2018). Effects of consistency between self and in-group on children's views of self, groups, and abilities. *Soc. Dev.* 27, 154–171. doi: 10.1111/sode.12255

Piatek-Jimenez, K., Cribbs, J., and Gill, N. (2018). College students' perceptions of gender stereotypes: making connections to the underrepresentation of women in STEM fields. *Int. J. Sci. Educ.* 40, 1432–1454. doi: 10.1080/09500693.2018.1482027

Pietrantoni, L., and Prati, G. (2008). Empowerment psicologico: contributo alla validazione italiana della scala di Spreitzer [Psychological empowerment: Contribution to the Italian validation of the Spreitzer scale]. *Risorsa Uomo* 14, 325–338. Available online at: https://www.francoangeli.it/riviste/Scheda_rivista.aspx?IDArticolo=34108

Riegle-Crumb, C., King, B., Grodsky, E., and Muller, C. (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. *Am. Educ. Res. J.* 49, 1048–1073. doi: 10.3102/0002831211435229

Rubiano-Matulevich, E., Hammond, A., Beegle, K., Kumaraswamy, S. K., and Rivera, S. (2019). *Improving the Pathway From School to STEM Careers for Girls and Women*. World Bank Blogs. Available online at: https://blogs.worldbank.org/opendata/improving-pathway-school-stem-careers-girls-and-women

Ruiz-Cantisani, M. I., Lopez-Ruiz, D. I., Suárez-Cavazos, N., Novelo-Villegas, J., Rincon-Flores, E. G., and Burgos-López, M. Y. (2021). "STEM & Gender equity: empowering women in vulnerable environments," in *2021 IEEE Global Engineering Education Conference (EDUCON)*. IEEE.

Salman, M. F., Olawoye, F. A., and Yahaya, L. A. (2011). Education reforms in Nigeria: Implications for the girl-child participation in sciences, technology and mathematics (STM). *Educ. Res. J.* 1, 1–8. Available online at:http://www.nigeria-education.org/content/education-reforms-nigeria-implications-girl-child-participation-sciences-technology-and

Salmela-Aro, K., and Upadyaya, K. (2012). The schoolwork engagement inventory: energy, dedication, and absorption (EDA). *Eur. J. Psychol. Assessment* 28, 60–67. doi: 10.1027/1015-5759/a000091

Schleicher, A. (2019). *Pisa 2018. Insights and Interpretations*. OECD. Available online at: https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf (accessed April 12, 2022).

Spreitzer, G. M. (1995). Individual empowerment in the workplace: Dimensions, measurement, and validation. *Acad. Manage. J.* 38, 1442–1465. doi: 10.2307/256865

Starr, C. R., and Simpkins, S. D. (2021). High school students' math and science gender stereotypes: relations with their STEM outcomes and socializers' stereotypes. *Soc. Psychol. Educ.* 24, 273–298. doi: 10.1007/s11218-021-09611-4

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *Am. Psychol.* 52, 613–629. doi: 10.1037/0003-066X.52.6.613

Tam, H. L., Chan, A. Y. F., and Lai, O. L. H. (2020). Gender stereotyping and STEM education: Girls' empowerment through effective ICT training in Hong Kong. *Child. Youth Serv. Rev.* 119:105624. doi: 10.1016/j.childyouth.2020.105624

Tandrayen-Ragoobur, V., and Gokulsing, D. (2021). Gender gap in STEM education and career choices: what matters? *J. Appl. Res. Higher Educ.* 14, 1021–1040. doi: 10.1108/JARHE-09-2019-0235

Thébaud, S., and Charles, M. (2018). Segregation, stereotypes, and STEM. *Soc. Sci.* 7:111. doi: 10.3390/socsci7070111

Tomasetto, C., Mirisola, A., Galdi, S., and Cadinu, M. (2015). Parents' math–gender stereotypes, children's self-perception of ability, and children's appraisal of parents' evaluations in 6-year-olds. *Contemp. Educ. Psychol.* 42, 186–198. doi: 10.1016/j.cedpsych.2015.06.007

UNESCO (2017). *Cracking the Code: Girls' and Women's Education in Science, Technology, Engineering and Mathematics (STEM)*. Available online at: https://unesdoc.unesco.org/ark:/48223/pf0000253479/PDF/253479eng.pdf.multi (accessed April 12, 2022).

UNICEF (2020a). *Mapping Gender Equality in STEM From School to Work*. Available online at: https://www.unicef.org/globalinsight/stories/mapping-gender-equality-stem-school-work (accessed April 10, 2022).

UNICEF (2020b). *Towards an Equal Future: Reimagining Girls' Education Through STEM*. Available online at: https://www.unicef.org/reports/reimagining-girls-education-through-stem-2020 (accessed April 10, 2022).

United Nations (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development*. A/RES/70/1. Available online at: https://sustainabledevelopment.un.org/post2015/transformingourworld (accessed April 10, 2022).

Van de Schoot, R., Lugtig, P., and Hox, J. (2012). A checklist for testing measurement invariance. *Eur. J. Dev. Psychol.* 9, 486–492. doi: 10.1080/17405629.2012.686740

van de Vijver, F. J. R., and Leung, K. (1997). *Methods and Data Analysis of Comparative Research*. Thousand Oaks, CA: SAGE.

Wang, M. T., and Eccles, J. S. (2012). Social support matters: Longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child Dev.* 83, 877–895. doi: 10.1111/j.1467-8624.2012.01745.x

Wilkinson, L. C., and Marrett, C. B. (1985). *Gender Influences in Classroom Interaction*. Orlando, FL: Academic Press.

World Bank (2014). *Higher Education in Science and Technology is Critical for Africa's Development*. Available online at: https://www.worldbank.org/en/news/feature/2014/03/12/higher-education-in-science-and-technology-is-critical-for-africas-development (accessed April 19, 2022).

World Bank (2022a). *Italy*. Available online at: https://data.worldbank.org/country/italy (accessed April 19, 2022).

World Bank (2022b). *Nigeria*. Available online at: https://data.worldbank.org/country/nigeria (accessed April 19, 2022).

World Economic Forum (2021). *Global gender gap. Report 2021*. Available online at: http://www3.weforum.org/docs/WEF_GGGR_2021.pdf

Wynn, A. T., and Correll, S. J. (2017). Gendered perceptions of cultural and skill alignment in technology companies. *Soc. Sci.* 6:45. doi: 10.3390/socsci6020045

Zilberman, A., and Ice, L. (2021). Why computer occupations are behind strong STEM employment growth in the 2019–29 decade. *Employ. Unemploy.* 10:1.

Zimmerman, M. A., and Warschausky, S. (1998). Empowerment theory for rehabilitation research: Conceptual and methodological issues. *Rehabil. Psychol.* 43, 3–16. doi: 10.1037/0090-5550.43.1.3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 03 November 2022
DOI 10.3389/fpsyg.2022.919870

OPEN ACCESS

EDITED BY
Joyce Endendijk,
Utrecht University, Netherlands

REVIEWED BY
Selma Korlat,
University of Vienna, Austria
Rachel Pauletti,
Lynn University, United States
Emil Smith,
Aarhus University, Denmark

*CORRESPONDENCE
William M. Bukowski
william.bukowski@concordia.ca

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 14 April 2022
ACCEPTED 12 July 2022
PUBLISHED 03 November 2022

CITATION
Kuzyk O, Gendron A, Lopez LS and Bukowski WM (2022) Gender and contextual variations in self-perceived cognitive competence.
Front. Psychol. 13:919870.
doi: 10.3389/fpsyg.2022.919870

Gender and contextual variations in self-perceived cognitive competence

Olivia Kuzyk[1], Alice Gendron[1], Luz Stella Lopez[2] and William M. Bukowski[1]*

[1]Department of Psychology, Concordia University, Montréal, QC, Canada, [2]Department of Education, Universidad del Norte, Barranquilla, Colombia

School performance and cognitive competence can be conceptualized as social and relational constructs. Thus, we expect their association to vary as a function of other socially-embedded variables which have proven meaningful in the academic domain. The present study takes a critical theory approach to assess gender-related and contextual variability in the association between peer-assessed school performance and self-perceived cognitive competence. The sample consisted of 719 preadolescents (M age = 9.5 years, range = 9 to 12.5 years) living in lower- and upper-middle-class neighborhoods in Montreal, Canada and Barranquilla, Columbia. Multigroup comparisons revealed that (a) peer-assessed school competence was more strongly associated with self-perceived cognitive competence for upper-middle-class than lower-middle-class participants from Barranquilla, whereas the opposite pattern was observed with Montreal participants, and (b) that the association between communal orientation and self-perceived cognitive competence was stronger for girls than for boys across the sample, especially in the upper-middle-class school in Montreal. These findings highlight the nuanced degree of gender differences in preadolescents' perceived academic competence and emphasize the role of SES in shaping self-perceptions.

KEYWORDS

cognitive competence, gender, childhood, culture, socioeconomic factors

Gender and contextual variations in self-perceived cognitive competence

The self is a dynamic construct that is shaped by experiences across the lifespan – especially in middle-childhood and adolescence (DuBois et al., 2000; Sebastian et al., 2008). Multiple proximal and distal factors influence the development of the self, including positive and negative aspects of individual experience (e.g., success and failure in achievement-related and social tasks), as well as other features of the school and peer environments (Bukowski and Raufelder, 2018). The present cross-cultural study

emphasizes the intersection between gender-related features at the level of the person and the contexts where the children are situated. The study examines the associations between variations in self-perceived cognitive competence, school performance, and aspects of gender in a sample of preadolescent girls and boys from upper- and lower-middle-class families in Montreal, Canada, and Barranquilla, Colombia. The goal of the study was to examine (a) how preadolescents' self-perceived cognitive competence is associated with school performance and with different aspects of gender and (b) how these associations vary as a function of cross-cultural context (i.e., place) and socioeconomic status (SES).

Self-perceived competence is defined as an individual's judgment of their own abilities, functioning, and well-being (Harter, 1996). Research on the self is typically guided by three premises (Harter, 2012); (1) that self-perceptions are only moderately associated with actual experiences, (2) that self-perceived competence can be affected by other person-related variables that can either increase or decrease a person's negative or positive self-views, and (3) that person-level and group-level variables can moderate the association between measures of functioning and self-perceptions. We used a broad multilevel perspective in our application of these premises to the study of the effects of gender and self-perceptions on cognitive competence. Our approach is characterized by two central ideas. The first is that we recognize that gender is a complex and multifaceted construct whose features need to be studied together to obtain a fuller view of how the defining aspects of gender work in concert to affect outcomes. Second, we maintain that gender is a social construct whose defining characteristics and given meanings are likely to vary across social and cultural contexts (World Health Organization [WHO], 2017). Incorporating these points related to the self and gender into our framework present some theoretical challenges, which are addressed in the following sections.

Gender theory

Our approach to these issues is inspired by critical theory (Bohman, 2021), and more specifically, by three fundamental claims from critical gender theory (Jule, 2014). The first claim is that simple comparisons between females and males provide only a very narrow assessment of the vast array of features and effects that constitute gender. In this study, we go beyond a simple binary comparison by including measures fashioned after the femininity and masculinity measures of the Bem Sex Role Inventory (BSRI) in order to capture gender-role traits that covary with cisgender measures of masculinity and femininity (Bem and Lewis, 1975). A second claim of critical gender theory is that to understand the dynamics of gender, one needs to assess how the facets of gender interface with actual experiences and social institutions.

In this study, we assessed how gender-related traits are associated with school performance. Instead of seeing school performance solely as a form of individual achievement, we also conceptualize it as a relational or participatory experience which may benefit from one's capacity to connect with the shared goals of the institutional environment. We see the gendered dimension of communal orientation as a trait that promotes effective functioning in academic tasks. We examined these factors by assessing how aspects of gender were related directly to self-perceived cognitive competence and how they moderated the association between school performance and self-perceived cognitive competence. The third claim is that gender is a social construct whose features and meanings vary across contexts. We assessed contextual variance by examining the effects of two broad contextual factors: the socioeconomic status of the children's school/neighborhood and place (i.e., whether participants were from Montreal. Canada, or Barranquilla, Colombia). We chose to study preadolescents from two cultural contexts that were likely to differ in their normative social relationship to gender and its multiple facets, as well as display differences in the gendered experiences they present to children in their respective settings when comparing lower- and upper-middle-class school environments. This decision is based on prior findings with classroom samples that support variance in gender identity as a function of SES (Bukowski et al., 2019, 2021). Our assessment of contextual factors focused on between-group differences in the degree to which associations between measures of gender (i.e., cisgender and gender roles) were associated with measures of self-perceived cognitive competence, and assessed whether gender moderated the association between school performance and self-perceived cognitive competence.

Self-perceived cognitive competence in an academic setting

Self-perceptions of academic competence are an understudied domain of research in relation to the self-concept. They are important because they affect subsequent goals in school tasks (Bong and Skaalvik, 2003; O'Mara et al., 2006). Children who endorse positive views of their cognitive competence have been shown to make more efforts to perform academically (Guay et al., 2003). Research also supports that academic attainment influences children's self-concept during a developmental period where academic self-perceptions are sensitive to experiences of success and failure (Skaalvik and Valås, 1999). As such, a bi-directional model (Marsh and Martin, 2011; Brunner et al., 2013) offers a more complete understanding of the reciprocal contributions between academic self-perceptions and performance.

There is a historical trend of gender differences in school performance throughout elementary school and into

adolescence (e.g., Brophy, 1985; Alexander et al., 1997; Dwyer and Johnson, 1997; Neuburger et al., 2012; Kingdon et al., 2017). Current evidence shows that the difference between girls and boys is relatively small (Voyer and Voyer, 2014) except perhaps in academic subjects that rely heavily on language skills (Reilly et al. (2019). There is some long-standing evidence that girls perceive themselves more positively in stereotypically feminine areas (i.e., reading and writing), but judge themselves more harshly on stereotypically masculine subjects (i.e., math and science) (Ruble et al., 1993). Moreover, girls evaluate themselves more negatively on measures of general self-worth (Kling et al., 1999) and report higher levels of school-related worry compared to boys (Silverman et al., 1995).

Explanations for this discrepancy may, in part, centre on differences on how boys and girls develop a sense of cognitive competence. Generally, researchers highlight that boys show higher scores on measures of self-worth compared to girls (e.g., Chubb et al., 1997; Quatman and Watson, 2001; Birndorf et al., 2005). This may be reflective of how boys and girls approach academics and manage evaluative feedback. That is, girls may regard these situations as opportunities to learn about their abilities and thus, may be more likely to internalize feedback (Roberts, 1991). These tendencies may motivate girls to do well, and also lead them correspondingly to experience more distress when they encounter failure or difficult feedback. Boys, conversely, are more competitive and may approach academics with more self-confidence and deny the evaluative feedback that is provided (Roberts, 1991). A self-confident approach may buffer the effects of failure or poor performance because it may lead boys to view feedback as less informative. This model signals potential variation in how children perceive their cognitive competence and perform in school based on the extent to which they ascribe to masculine and feminine traits.

Therefore, one can speculate that the processes of academic achievement can be conceptualized in relation to forms of functioning that are differentially associated with feminine and masculine gender roles. The items assigned to the feminine and masculine scale of the Bem Sex Role Inventory can be conceived of as fitting the well-established dimensions of communion and agency that are known to be gendered aspects of functioning (Abele and Wojciszke, 2007; Abele et al., 2016). Whereas the personal features of instrumentality and assertiveness associated with the masculine gender role (Bem, 1974) support a view of academic competence as a form of individual achievement that results from personal action, the personal feature of communion associated with the feminine gender role may support a view of competent school functioning as a collective activity that requires a commitment to group processes, which derive from group-sanctioned forms of knowledge. Hence, one could expect communion to promote self-perceived academic competence to the degree that functioning in school rests on participation in communal activities. According to this perspective, adhering to gender roles and perceiving oneself as cognitively competent may be overlapping forms of self-perception. As a result, boys who see themselves as being assertive and instrumentally competent may also see themselves as academically competent, as this form of competence is an expression of their personal assertiveness and instrumental skills. Similarly, girls who see themselves as communal may also see themselves as academically competent as this form of competence is an expression of their capacity to function in a domain that requires communal skills. This perspective is supported by evidence that adolescents perceive schools to be more feminine than masculine (Heyder and Kessels, 2013). Based on this reasoning, one can hypothesize that (a) gender roles may be univariate correlates of measures of self-perceived cognitive competence and (b) they may function as moderators that strengthen the association between achievement and self-perceived cognitive competence. Each of these hypotheses will be examined in our analyses.

Self in context

Socioeconomic status and culture are likely to affect variations in the self-concept during preadolescence. Children are situated within rich networks of influence, and thus it is unsurprising that these contexts impact their self-worth and academic achievement. Children belonging to high-SES families report higher levels of self-worth compared to those from low-SES backgrounds (Rhodes et al., 2004). However, this relationship is dependent on the importance placed on academic achievement. Campbell et al. (2002) have reported that this pattern has been observed as a result of a stronger emphasis placed on academic achievement in high-SES families.

Similarly, Santo et al. (2013) found that cognitive competence was more strongly related to self-worth in a sample of early adolescents from a low-SES background, whereas social competence was more closely linked to self-worth among children in high-SES groups. The authors posited that these patterns reflect differences in culturally-determined indicators of self-worth. For low-SES peer groups, pursuing an education may be a strong indicator of success, whereas achieving or maintaining social status may be more important for higher-SES families. Divergent patterns were also observed for adolescents from individualistic cultures relative to collectivistic cultures. More specifically, the association between perceived cognitive competence and self-worth was weaker for those from a collectivistic society compared to those from an individualistic one. Cognitive competence may be more closely aligned with individualistic values and thus regarded as less important to collectivistic groups (Santo et al., 2013). There is no doubt that differences in how early adolescents perceive their cognitive competence and general worth are complex and salient across cultural groups.

The current study

Broadly, the focus of the present study was to examine contextual variations in how young adolescents' self-perceived cognitive competence is associated with their academic achievement. Here, gender is the primary contextual variable of interest, in that the current research builds on historical trends of gender differences in scholastic performance and provides objective indices of how they relate to cognitive competence. We conceptualize gender as a multidimensional construct. The current research examined the extent to which self-assessed cognitive competence and peer-assessed academic performance varies as a function of both masculine and feminine features of gender. It was hypothesized that children who identify more strongly with feminine features will demonstrate a stronger association between perceived cognitive competence and academic achievement.

Moreover, in light of research emphasizing the importance and complexity of the cultural context, we aimed to examine how this relationship changes across interactions of SES and cultural groups. We proposed that for children who identify with feminine traits, the relationship between their self-perceptions of cognitive competence and academic achievement will be strongest in the low-SES individualistic group. Second, we hypothesize a significant association between cognitive competence and academic achievement for children who reported more feminine traits in the low-SES collectivistic group.

Methods

Participants

The sample consisted of 719 (M age = 10.70 years, SD = 1.20) fourth-, fifth- and sixth-grade girls (N = 380) and boys (N = 339) in mixed-sex schools located in lower-middle- and upper-middle-class neighborhoods in Montreal, Canada (N = 302) and Barranquilla, Colombia (N = 417). The proportion of boys and girls, and of upper- and lower-middle-class participants, was roughly the same in each country. Socioeconomic status was operationally defined with different criteria for the two places. In Colombia, this designation was based on an index of neighborhood SES known as *estrato* that is assigned by the Colombian government based on the quality of housing and services in the neighborhood (Rueda-García, 2003). Scores range from 1 to 6, with higher scores indicating greater affluence. The mean estrato score for the children from lower-middle-SES schools was 2.52, (SD = 0.70) indicating that the participants at the low-SES schools were indeed within the lower socioeconomic strata. Although individual estrato ratings were not obtained from the high-SES school sampled in Barranquilla, school officials indicated that children who attended this school typically fell into the highest estrato category (6). The data were collected in 2002.

SES for the Montreal children was based on the average family income of children in their school. Parents completed a questionnaire in which they selected the income level (from 10 choices ranging from below $15,000 to over $95,000) that was closest to that of each adult member of the household in the last year. A total income score was calculated by adding the income of each family member. There were large between-school differences: one school had a mean family income of $36,027 CND, a second school had a mean of $68,400 and the third school had a mean of $79,194. The first school was designated as lower-middle class and the second two schools as upper-middle class. Based on information from the 2001 Canadian census (the census conducted closest to the time of the data collection), the mean family income of participants from the first school was considerably lower than the provincial average of $59,296, whereas the mean family income of participants in the latter two schools was above the provincial average (Statistics Canada, 2002). In the Barranquilla part of the sample, there were 149 participants from the two schools in lower- middle-class neighborhoods and 268 participants from the one school whose students came from upper-middle-class neighborhoods. In the Montréal part of the sample, there were 149 participants from the one school in a lower- middle-class neighborhoods and 268 participants from the two schools in upper-middle-class neighborhoods.

Procedure

A multi-stage recruitment process was used in each city. In Montreal, permission was first obtained from the relevant school board, and then from school principals. Active consent was required from parents of potential participants. In Barranquilla, the parents of the potential participants were informed by the school principal of the purposes and procedures of the study. They were also informed that participation in the study was voluntary. Parents could ask for their child not to be included in the study. In this region of Colombia, school principals often act *in loco parentis*. Their rights as participants were explained to them prior to the beginning of the data collection. Each participating child provided assent to be in the study. Using these recruitment procedures, a participation rate of over 85% was obtained in Montreal and of over 90% in Barranquilla.

The children completed a questionnaire at their desks in their classrooms in a group administration. The Colombian participants completed a version of the questionnaire that had been translated into Spanish by translators working in the areas of education and psychology. This adaptation was also backtranslated into English by a separate group of translators to ensure that the meaning of items was retained in the process.

Measures

The participants completed three measures: (a) a peer assessment measure of school performance, (b) an altered version of Harter's (1982) Perceived Competence Scale for Children, and (c) an abbreviated version of the Bem Sex Role Inventory Bem (1974). The participants completed these inventories *via* a paper-and-pencil format at their desks in class. At least three members of the project team were in each classroom to make sure the participants understood the instructions and to answer any questions about how to complete the measures.

Peer assessment measure

Peer assessment procedures are used to assess how children are perceived by their peers. These procedures are known to provide valid and reliable measures of children's competence and effective functioning (Bukowski et al., 2012). In a peer assessment procedure, participants are shown a list of items that describe forms of functioning and are asked to indicate which of their participating classmates fit each description. In this study, two items were used to assess school performance. They were "Someone who is smart and does well in school" ("Es intelligente y tiene un buen rendimiento academico en le escuela") and "Someone who always knows the right answers in school" ("Siempre sabe la respuesta correcta en la escuela"). Two scores were calculated for each participant on each item. These values correspond to the number of times the child was nominated for the item by same- and other-gender peers. Each score for each item was adjusted for possible biases that may result from variations in group size (see Velásquez et al., 2013). Separate corrections were made for the same-gender and other-gender measures. In this study, only the same-gender measures were used. A school performance score was computed for each participant by adding the two class-size-adjusted same-gender scores together. When assessed using Cronbach's alpha, the reliability of this aggregated score was observed to be 0.92. The use of a peer assessment measure is advantageous as it provides a common measurement procedure across the schools and contexts included in the study. Other forms of measurement, such as school grades, can be problematic due to variations in the procedures used in different schools and places. The mean and standard deviation for this measure are 1.22 and 1.88.

Bem sex role inventory

The participants rated ten words taken from Bem's (1974) BSRI. Two of these words were "feminine" and "masculine;" the other eight words were chosen based on two criteria. First, we chose words for which there was strong empirical evidence of their alignment with the femininity and masculinity dimensions used in Bem's (1974) initial studies. Second, the words had to be relevant to the preadolescent participants in the study. The four words were taken from the femininity scale were "Affectionate," "Sympathetic," "Understanding," and "Sensitive to the Needs of Others." This set of items was seen as representative of communal orientation. The four words taken from the masculinity dimension were "Independent," "Athletic," "Leader" and "Forceful." They were interpreted as representing instrumentality/assertiveness. Using a five-point scale in which a "1" represented "Not like me at all" ("No me describe") and a "5" equated to "Just like me" ("Me describe"), each participant rated each word according to whether it provided a true description of the self. The scores on the items for each measure were initially analyzed with a principal components factor analysis. The observed factor loadings were used to create a communal orientation score and an instrumentality/assertiveness score for each participant. To create these scores, the items were weighted by the observed factor loadings from the PCA. The internal consistency of these scales, assessed with omega, was 0.77 and 0.82 for the instrumentality/assertiveness and communal orientation scales, respectively. The mean and standard deviation for the instrumentality/assertiveness measure are 3.69 and 1.00; the mean and standard deviation for the communal orientation measure are 3.93 and 0.97.

Perceived competence scale for children

Self-perceived cognitive competence was measured using selected items from Harter's (1982) Perceived Competence Scale for Children. A set of seven items adapted from Harter's original scale were used to assess positive views of cognitive competence. Consistent with the rating scale concerns raised by Yeager and Krosnick (2011), the items were written to fit a simple five-point scale in which 1 meant "never true" and 5 meant "always true." The preadolescents were instructed to read each description and indicate how well each one fit their self-view. The items were "I feel that I am very good at my school," "I feel like I am just as smart as other kids my age," "I like school because I do well in school," "I am pretty slow in finishing my schoolwork," "I often forget what I learn," "I wish it were easier to understand what I read," and "I have trouble figuring out the answers in school." The last four items were reversed self-perceived competence items. As with the procedures used with the BSRI items, the scores on these items were initially analyzed with a principal components factor analysis. The observed factor loadings were used to create a self-perceived competence score for each participant. To create this score, the items were weighted by the observed factor loading from the PCA. The internal consistency of this scale, assessed with omega, was observed to be 0.76. The

mean and standard deviation for this measure are 3.65 and 0.79. (The means (and standard deviations) for all the person-level variables are shown in Table 1 for the categorical combinations of cisgender, SES, and place.)

Other variables included in the analyses were place, coded as −1 for Montreal and 1 for Barranquilla, cisgender (i.e., the gender assigned to the child at birth) coded as −1 for boys and 1 for girls, and SES coded as −1 for lower-middle-class and 1 for upper-middle-class.

Results

Analyses were conducted with Mplus (Muthén and Muthén, 2015). A two-phase procedure followed. In the first phase, person-level variables were used as predictors of the outcome measure (i.e., the measure of self-perceived cognitive competence). In the second phase, mulitigroup comparisons were performed to assess whether any of the associations observed in the first phase differed as a function of place (i.e., Barranquilla and Montreal), SES, and the intersection between place and SES.

In the first phase, eleven variables were used as predictors of the dependent variable (i.e., self-perceived cognitive competence). These predictors were used to capture the univariate and the interactive effects of the peer-assessed measure of academic performance and the three gender measures. The eleven predictors were: (a) the peer-assessed measure of school performance, (b) the participant's cisgender, (c) the measure of instrumentality/assertiveness, (d) the measure of communal orientation, (e) the two-way interaction between the peer-assessed measure of school performance and the cisgender measure, (f) the two-way interaction between peer-assessed school performance and communal orientation, (g) the two-way interaction between peer-assessed school performance and the measure of instrumentality/assertiveness, (h) the two-way interaction between the cisgender measure and communal orientation, (i) the two-way interaction between the cisgender measure and the measure of instrumentality/assertiveness, (j) the three-way interaction between the cisgender measure, peer-assessed school performance, and communal orientation, and (k) the three-way interaction between the cisgender measure, peer-assessed school performance, and instrumentality/assertiveness. The statistically significant findings are reported in Table 2.

Initial analyses revealed statistically significant coefficients for five of the predictors, specifically (a) the peer-assessed school competence measure (standardized coefficient = 0.36, standard error = 0.03 t = 10.90, p < 0. 001), (b) cisgender (standardized coefficient = 0.08, standard error = 0.04 t = 2.33, p < 0.02), (c) communal orientation (standardized coefficient = 0.12, standard error = 0.04, t = 3.38, p < 0.001), (d) the two-way interaction between the peer-assessed school performance measure and cisgender (standardized coefficient = –0.10, standard error = 0.035, t = –2.87, p < 0.005), and (e) the two-way interaction between communal orientation and cisgender (standardized coefficient = 0.071, standard error = 0.039, t = 2.44, p < 0.15). A clarification of the two-way interaction between the peer-assessed school performance measure and cisgender indicated that the association between self-perceived cognitive competence and peer-assessed school performance was stronger for boys (coefficient = 0.43) than girls (coefficient = 0.27). A clarification of the two-way interaction between communal orientation and cisgender indicated that the association between self-perceived cognitive competence and

TABLE 1 Means (standard deviations) for person-level variables for groups defined by participant gender, SES, and place.

Group	Self-perceived cognitive competence	Peer-assessed school performance	Communal orientation	Instrumentality/ Assertiveness
Boys, Barranquilla, Lower Middle Class	3.40 (0.72)	0.84 (1.41)	3.92 (0.88)	3.92 (0.94)
Girls, Barranquilla, Lower Middle Class	3.53 (0.79)	1.01 (1.49)	4.11 (0.81)	4.00 (0.84)
Boys, Montréal, Lower Middle Class	3.75 (0.91)	1.36 (1.69)	3.91 (0.84)	3.99 (0.93)
Girls, Montréal, Lower Middle Class	3.96 (0.76)	1.45 (1.70)	4.34 (0.52)	3.66 (0.72)
Boys, Barranquilla, Upper Middle Class	3.62 (0.75)	1.22 (1.99)	3.63 (1.18)	3.60 (1.20)
Girls, Barranquilla, Upper Middle Class	3.70 (0.79)	1.09 (1.86)	3.93 (1.21)	3.47 (1.04)
Boys, Montréal, Upper Middle Class	3.51 (0.77)	1.27 (1.72)	3.77 (0.87)	3.76 (1.03)
Girls, Montréal, Upper Middle Class	3.69 (0.78)	1.30 (1.65)	3.96 (0.75)	3.54 (0.97)

TABLE 2 Person-related predictors of the self-perceived cognitive competence score.

Level 1 variable	Standardized coefficients (standard errors)	*t* score (P-value)
Peer Measure	0.36 (0.03)	10.90 (0.001)
Cisgender	0.08 (0.03)	2.33 (0.02)
Communal Orientation	0.12 (0.04)	3.38 (0.001)
Peer Measure by Cisgender	−0.10 (0.04)	−2.87 (0.005)
Cisgender by Communal Orientation	0.07 (0.04)	2.05 ($p < 0.05$)

communal orientation was stronger for girls (coefficient = 0.19) than for boys (coefficient = 0.04).

Multigroup comparisons, conducted with Mplus, were then performed to assess whether these associations differed (a) for the participants from the two places, (b) for the participants from the lower-middle-class and upper-middle-class schools, and (c) for the participants from the four groups defined by a combination of place and SES (i.e., lower-middle-class participants from Barranquilla, lower-middle-class participants from Montreal, upper-middle-class participants from Barranquilla, and upper-middle-class participants from Montreal). Each multigroup comparison consisted of a two-step process (see Wang and Wang, 2019). In the first step, equality constraints were used to set the coefficients for a particular association to be equal across groups (e.g., the upper-middle-class and the lower-middle-class participants). If the coefficients for these groups were equal to each other, then setting them to be equal would not affect the overall fit of the model. If the coefficients were not equal to each other, then setting them to be equal would have an adverse effect of model fit. This negative effect of model fit would be manifested in an increase in the Chi-square value. In the second step of this comparative procedure, a chi-square difference test was used to assess the statistical significance of the change in the chi-square value.

Comparisons of the coefficients observed with the participants from the two places revealed no statistically significant differences. Comparisons that assessed differences between the participants from lower-middle-class and upper-middle-class schools revealed only one statistically significant difference. Specifically, the two-way interaction between cisgender and communal orientation was observed to be weaker and statistically non-significant with the participants from the lower-middle-class schools (standardized coefficient = –0.03, standard error = 0.058, $t = -0.53$, $p > 0.5$), whereas it was statistically significant with the participants from the upper-middle-class schools (standardized coefficient = 0.13, standard error = 0.044, $t = 2.88$, $p < 0.005$). The positive coefficient observed with this two-way interaction for the participants from the upper-middle-class schools indicates that the effect of a communal orientation was stronger for girls than for boys.

Multigroup comparisons conducted with the four groups defined by a combination of place and SES revealed three between-group differences. First, the measure of peer-assessed school competence was observed to be more strongly associated with the outcome measure for the upper-middle-class participants from Barranquilla (standardized coefficient = 0.41) than for the lower-middle-class participants from Barranquilla (standardized coefficient = 0.26). The corresponding values for the upper-middle-class and lower-middle-class participants from Montreal were 0.28 and 0.36, respectively. These coefficients did not differ from each other. All of these coefficients were statistically significant. It is important to note that the differences between the upper-middle-class participants and lower-middle-class participants showed a different pattern in Montreal (lower-middle class was higher than upper-middle class) than in Barranquilla (lower-middle class was lower than upper-middle class).

A second difference was observed with the association between communal orientation and self-perceived academic competence. This association was observed to be more stronger for the lower-middle-class participants from Montreal (standardized coefficient = 0.19, standard error = 0.082, $t = 2.25$, $p < 0.02$) than for the lower-middle-class participants from Barranquilla (standardized coefficient = −0.024, standard error = 0.082, $t = 0.29$, $p < 0.75$). The corresponding values for the upper-middle-class participants from Montreal and Barranquilla were 0.07 (standard error = 0.075, $t = 0.96$, $p > 0.3$) and 0.14 (standard error = 0.055, $t = 2.63$, $p < 0.009$). Again, a different pattern of findings was observed in Montreal (lower-middle class was higher than upper-middle class) than in Barranquilla (lower-middle class was lower than upper-middle class).

The third set of differences was observed with the association between two-way interaction between cisgender and communal orientation and the measure of self-perceived academic competence. The coefficients for the association between this interaction score and the outcome were observed to be positive and statistically significant with the participants from upper-middle-class schools in Montreal (standardized coefficient = 0.14, standard error = 0.07, $t = 1.98$, $p < 0.05$) and Barranquilla (standardized coefficient = 0.12, standard error = 0.055, $t = 2.15$, $p < 0.03$) and negative and statistically non-significant with the participants from lower-middle-class schools in Montreal (standardized coefficient = –0.09, standard error = 0.082, $t = -1.29$, $p > 0.3$) and Barranquilla (standardized coefficient = –0.022, standard error = 0.082, $t = -0.26$, $p > 0.7$). Group comparisons indicated that the coefficients for the participants from the upper-middle-class school differed from the coefficient observed with the participants from the lower-middle-class schools in Montreal. The positive value of the coefficients observed with the participants from the

upper-middle-class school indicates that for these participants, the association between communal orientation and the outcome measure (i.e., the measure of self-perceived cognitive competence) is stronger for girls than for boys.

Discussion

Two key findings were revealed. The first is that the measures of gender roles are associated with self-perceived cognitive competence as univariate predictors and as moderators. As importantly, our findings were varied as a function of place and SES. These findings point to the complex pattern of the factors associated with self-perceived cognitive competence and its association with specific components of gender. The findings confirm two basic features of the study's conceptual frame. Specifically, the findings show that the associations observed with gender-related variables will vary as a function of contextual factors – especially intersection between culture and SES. The findings also show that adherence to gender roles is associated with self-perceived cognitive competence in a direct manner and as a moderator of experience. This evidence of the importance of gender role adherence was, however, observed only with the dimension of communal orientation and only in particular contexts.

A primary finding from the study is the observation that the association between peer-assessed school performance and self-perceived cognitive competence is moderated by the cisgender measure, and that this interaction is moderated by an interaction between place and SES and by the cisgender measure. Peer-assessed school competence was observed to be more strongly associated with the outcome measure for the upper-middle-class participants from Barranquilla than for the lower-middle-class participants from Barranquilla. The opposite pattern was observed with the Montreal participants; albeit to a smaller and statistically non-significant degree. The moderating effect of the cisgender measure indicated that the association between peer-assessed school performance and self-perceived cognitive competence was weaker for girls than for boys. Consistent with prior findings, self-perceptions of cognitive competence appear to be less dependent on actual experience for boys than for girls. These findings confirm our prior results, observed with a different sample, that gender differences may be stronger for upper-middle-class children in the Colombian context (Santo et al., 2013). They also provide an explanation for Van Houtte's (2004) observation of stronger achievement levels among boys than girls.

The second important result pattern also points to a difference between girls and boys. This two-way interaction indicates that the association between communal orientation and self-perceived cognitive competence was stronger for girls than for boys. This finding provides partial support for our reasoning that the gender-role measure may overlap with the self-perceived competence measure. To a small degree, girls who see themselves as communally oriented also see themselves and being competent in cognitive tasks. This pattern was further moderated by SES and place, and was seen only among the participants from the upper-middle-class school in Montreal. Hence, this shows that the effect of gender varies as a function of culture and SES. In this way, these findings provide partial support for our speculation that gender roles are intertwined with perceptions of cognitive competence. This evidence was, however, limited in two ways. First, it was observed only with the measure communal orientation. Second the effect of communal orientation was observed only for girls from upper-middle-class neighborhoods in the two places. These findings reveal a high level of specificity in gender-related findings. Together, these findings emphasize the importance of gender in models of self-perceived academic competence.

Our analyses revealed three statistically significant univariate findings and two statistically significant two-way findings at the level of the person. Positive associations were observed between self-perceived cognitive competence and (a) peer-assessed school performance, (b) the cisgender measure (i.e., girls showed stronger judgments of their cognitive abilities than boys) and (c) communal orientation. Additionally, statistical analyses involving two-way interactions across these measures (i.e., cisgender and school performance; cisgender and communal orientation) revealed that academic achievement was more predictive of boys' self-judgments of their cognitive competence as compared to girls, whereas communal orientation was more predictive of self-perceived cognitive competence for girls.

The group comparisons show that the meaning of gender around scholastic achievement and self-assessed cognitive competence is contextually dependent, particularly across SES groups. That is, a two-way interaction between cisgender and communal orientation was predictive of cognitive competence among children attending upper-middle-class schools, but not lower-middle-class schools. Notably, this effect appeared stronger for girls relative to boys. Between-group comparisons further highlighted the complexities of contextual variations, in that communal orientation was related to self-assessed cognitive competence in lower-middle-class schools in Montreal, but not Barranquilla. Additionally, the interaction between cisgender and communal orientation was predictive of the outcome in upper-middle class schools in both Montreal and Barranquilla. Analyses also revealed that this effect was stronger for girls than it was for boys. Perhaps the most important finding from the study is the observation that the self-perceptions of cognitive competence among girls and boys from lower SES neighborhoods in Barranquilla appear to be unaffected gender roles. This finding is important as it supports the basic premise of the study that the significance of gender varies across cultural contexts. Although an exact interpretation of this pattern of findings is elusive, at the very least they indicate that the

meaning of the measures of gender used in our analyses are different for the low SES participants from Barranquilla. It may be that the concepts themselves (i.e., a communal orientation and assertiveness/instrumentality) are not as "gendered" for the low SES participants from Barranquilla. A further exploration of these findings may benefit from an assessment of how these measures are associated with gender-related constructs such as gender typicality and felt pressure to conform (see Egan and Perry, 2001) to conform and whether these associations vary as a function of SES and culture. Together, these findings indicate that the SES composition of classrooms across geographic location shape the gender norms around academic achievement and cognitive competence.

Researchers have already suggested that hegemonic gender norms are evoked and sanctioned depending on the social context in which they occur (Ridgeway and Correll, 2004; Morris, 2011; Hsin, 2018). The school context is therefore a major channel for how these gender norms are expressed and actualized in young adolescents' achievement outcomes (Hsin, 2018). Ethnographic studies demonstrate that divergent achievement patterns for boys and girls evolve from cultures of masculinity that minimize the importance of boys performing well academically (e.g., United Kingdom: Mac an Ghaill, 1994; Australia: Martino, 1999; United States: Pascoe, 2007). Certain academic disciplines and study behaviors are regarded as "feminine," which has been shown to negatively affect boys' motivation toward school (Pajares and Valiante, 2001; Bhanot and Jovanovic, 2005). In fact, traits linked to femininity as well as those that are consistent with studious attitudes (e.g., being tidy, cooperative and passive) may even be advantageous for girls (Jones and Myhill, 2004; Beaman et al., 2006). Our findings support this view. Other research has shown that boys' peer groups in secondary school have tendencies to engage in less studious behaviors compared with girls, which notably accounts for the lower academic performance observed in boys (Van Houtte, 2004). Our findings fit well within this body of work. They highlight that the extent to which young adolescents, particularly girls, who identify with features of communal orientation also hold self-perceptions of their cognitive competence. A communal orientation may be largely consistent with the studious behaviors that have been identified in previous studies.

Furthermore, research has supported that gender differences in academics are strongly impacted by the SES composition of schools (Legewie and DiPrete, 2014). Legewie and DiPrete (2014) reported that high-SES classrooms promote girls' academic achievement because they are not gendered as "feminine" in terms of interests and pursuits. Interestingly, they also encourage boys' educational outcomes by influencing their choice in the science, technology, engineering and mathematics (STEM) fields. Qualitative studies further demonstrate how hegemonic gender expectations are promoted through school environments—particularly by emphasizing engagement in sports culture over academics for boys (Morris, 2008). Participating in sports reflects an expression of hegemonic masculinity by demonstrating toughness and physical strength (Morris, 2008). Taken together, schools with a high-SES compositions do not regard academic achievement as a feminine pursuit, but more subtly enforce gendered behavior and interests for boys through the promotion of STEM trajectories and sports engagement.

The present set of findings are also consistent with this view but may reflect differences in the extent to which gender expectations in academic achievement are actualized across the developmental trajectory. We observed that the gender differences on self-competence were strongest in upper-middle-class schools in Montreal and Barranquilla. This may signal to more salient gender norms around academics for girls in upper-middle-class SES compositions. Similarly, it is also possible that there is less emphasis on STEM trajectories and sports engagement for boys in early adolescence, and thus less opportunity to shape self-perceived cognitive competence. Research in the field of self-efficacy (i.e., individuals' judgments around their ability to engage in behaviors that are required to achieve a desired objective) demonstrate gender differences emerging in early adolescence which increase over development (Huang, 2013). As such, we would expect divergent patterns to emerge as boys become exposed to more specific pressures for gender conformity as they progress in the school system and make choices about their future vocation. Specific academic courses were found to be important moderators of self-efficacy, in that previous work identified boys as having higher self-efficacy scores in mathematics and computer sciences, whereas girls showed elevations in language arts and small advantages on general academic self-efficacy (Van Houtte, 2004). Therefore, it would be of benefit for researchers to use statistical network analyses to examine how features of gender and cognitive competence vary as a function of academic courses, in addition to SES school composition and place variables.

Some limitations should be noted. First, the use of a cross-sectional design prevented causal interpretations. Follow-up studies using longitudinal designs are needed. Second, although multiple measure of gender were used, one can imagine that including more measures of gender identity would add diversity to the findings. Third, the data were collected 20 years ago. Given that some aspects of gender identity may have changed in the intervening years (Donnelly and Twenge, 2017), a replication study using more recent data is needed. Fourth, the study relies to a great extent on self-report measures. Aspects of gender and the dimensions of the self-concept are typically with self-report procedures. The use of peer reports might add to the currently available measures of gender. Fifth, SES is a multilevel concept (Bukowski et al., 2020). Although it is often measured as a feature of an individual or a family, SES was

used here ad a measure of the school context. A more complex approach to SES that included measure at the level of the group and the individual would expand our understanding of how SES intersects with gender and self perceptions of competence. Sixth, a richer conceptualization of gender is needed to understand the degree to which gender identity should be conceived of as a trait or as a conscious form of self-perception related to one's gender.

In conclusion, the present set of findings builds upon existing research to provide further insight into gender-related variations in self and academic achievement in early adolescence and across socio-geographical contexts. Our work highlights the specificity of gender differences in self-perceived cognitive competence in upper-SES compositions in schools, and thus ascribes meaning to features of gender that are dependent on gender expectations for scholastic achievement. While this study helps explain contextual variations in how young adolescents' self-perceptions of their cognitive competence are associated with their academic achievement, further research is required to disentangle course-specific nuances in order to reduce gender gaps and promote equality in academic achievement.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Concordia University Human Research Ethics Committee approved the study and it conforms to the recognized standards by the Declaration of Helsinki. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Funding

OK and WB were supported by the Social Sciences and Humanities Research Council of Canada and the Fonds de recherche du Queìbec: Socieìteì et culture.

Acknowledgments

We thank the schools, teachers, principals, parents, participants, and all those who helped with the data collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., and Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness—Communion with warmth and morality. *Frontiers in psychology* 7:1810. doi: 10.3389/fpsyg.2016.01810

Abele, A. E., and Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of personality and social psychology* 93, 751. doi: 10.1037/0022-3514.93.5.751

Alexander, K., Entwisle, D., and Horsey, C. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education* 70, 87–107. doi: 10.2307/2673158

Beaman, R., Wheldall, K., and Kemp, C. (2006). Differential teacher attention to boys and girls in theclassroom. *Educational review* 58, 339–366. doi: 10.1080/00131910600748406

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of consulting and clinical psychology* 42, 155. doi: 10.1037/h0036215

Bem, S. L., and Lewis, S. A. (1975). Sex role adaptability: One consequence of psychological androgyny. *Journal of personality and Social Psychology* 31, 634. doi: 10.1037/h0077098

Bhanot, R., and Jovanovic, J. (2005). Do parents' academic gender stereotypes influence whether they intrude on their children's homework? *Sex roles* 52, 597–607. doi: 10.1007/s11199-005-3728-4

Birndorf, S., Ryan, S., Auinger, P., and Aten, M. (2005). High self-esteem among adolescents: Longitudinal trends, sex differences, and protective factors. *Journal of Adolescent Health* 37, 194–201. doi: 10.1016/j.jadohealth.2004.08.012

Bohman, J. (2021). "Critical Theory," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Stanford, CA: Stanford University). (Spring 2021 Edition)*y.

Bong, M., and Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational psychology review* 15, 1–40. doi: 10.1023/A:1021302408382

Brophy, J. (1985). "Interactions of male and female students with male and female teachers," in *Gender influences in classroom interaction*, 115–142. *. doi: 10.1016/B978-0-12-752075-9.50011-8

Brunner, M., Gogol, K. M., Sonnleitner, P., Keller, U., Krauss, S., and Preckel, F. (2013). Gender differences in the mean level, variability, and profile shape of student achievement: Results from 41 countries. *Intelligence* 41, 378–395. doi: 10.1016/j.intell.2013.05.009

Bukowski, W. M., Castellanos, M., Commisso, M., Persram, R., and Lopez, L. S. (2019). Contextual variations in children's perceptions of their peers as gender typical. *International Journal of Behavioral Development* 43, 309–313. doi: 10.1177/0165025419846427

Bukowski, W. M., Cillessen, A. H. N., and Velasquez, A. M. (2012). "The use of peer ratings in developmental research," in *Handbook of developmental research methods*, eds B. Laursen, T. Little, and N. Card (New York, NY: Guilford Press), 211–228.

Bukowski, W. M., Dirks, M., Persram, R. J., Wright, L., and Infantino, E. (2020). Peer relations and socioeconomic status and inequality. *New Directions for Child and Adolescent Development* 2020, 27–37. doi: 10.1002/cad.20381

Bukowski, W. M., Dirks, M., Persram, R., Santo, J. B., DeLay, D., and Lopez, L. (2021). Contextual variations in associations between measures of aggression and withdrawal and functioning with peers: A replication study. *Developmental Psychology* 57, 2022–2031. doi: 10.1037/dev0000958

Bukowski, W. M., and Raufelder, D. (2018). "Peers and the self," in *Handbook of peer interactions, relationships, and groups*, eds W. M. Bukowski, B. Laursen, and K. H. Rubin (New York, NY: The Guilford Press), 141–156.

Campbell, F. A., Pungello, E. P., and Miller-Johnson, S. (2002). The development of perceived scholastic competence and global self-worth in African American adolescents from low-income families: The roles of family factors, early educational intervention, and academic experience. *Journal of Adolescent Research* 17, 277–302. doi: 10.1177/0743558402173004

Chubb, N. H., Fertman, C. I., and Ross, J. L. (1997). Adolescent self-esteem and locus of control: A longitudinal study of gender and age differences. *Adolescence* 32, 113.

Donnelly, K., and Twenge, J. M. (2017). Masculine and feminine traits on the Bem Sex-Role Inventory, 1993–2012: A cross-temporal meta-analysis. *Sex Roles: A Journal of Research* 76, 556–565. doi: 10.1007/s11199-016-0625-y

DuBois, D. L., Tevendale, H. D., Burk-Braxton, C., Swenson, L. P., and Hardesty, J. L. (2000). Self-system influences during early adolescence: Investigation of an integrative model. *The Journal of Early Adolescence* 20, 12–43. doi: 10.1177/0272431600020001002

Dwyer, C. A., and Johnson, L. M. (1997). "Grades, accomplishments, and correlates," in *Gender and fair assessment*, eds W. W. Willingham and N. S. Cole (Hillsdale, NJ: Lawrence Erlbaum Associates Publishers), 127–156.

Egan, S. K., and Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology* 37, 451–463. doi: 10.1037/0012-1649.37.4.451

Guay, F., Marsh, H. W., and Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of educational psychology* 95, 124. doi: 10.1037/0022-0663.95.1.124

Harter, S. (1982). The perceived competence scale for children. *Child Development* 53, 87–97. doi: 10.2307/1129640

Harter, S. (1996). "Teacher and classmate influences on scholastic motivation, self-esteem, and level of voice in adolescents," in *Social motivation: Understanding children's school adjustment*, eds J. Juvonen and K. R. Wentzel (Cambridge: Cambridge University Press), 11–42. doi: 10.1017/CBO9780511571190.004

Harter, S. (2012). *Self-perception profile for adolescents: Manual and questionnaires*. Denver, CO: University of Denver. doi: 10.1037/t05703-000

Heyder, A., and Kessels, U. (2013). Is school feminine? Implicit gender stereotyping of school as a predictor of academic achievement. *Sex Roles: A Journal of Research* 69, 605–617. doi: 10.1007/s11199-013-0309-9

Hsin, A. (2018). Hegemonic gender norms and the gender gap in achievement: The case of Asian Americans. *Sociological Science* 5, 752–774. doi: 10.15195/v5.a32

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education* 28, 1–35. doi: 10.1007/s10212-011-0097-y

Jones, S., and Myhill, D. (2004). 'Troublesome boys' and 'compliant girls': Gender identity and perceptions of achievement and underachievement. *British Journal of Sociology of Education* 25, 547–561. doi: 10.1080/0142569042000252044

Jule, A. (2014). "Gender theory," in *Encyclopedia of Quality of Life and Well-Being Research*, 2464–2466. *. doi: 10.1007/978-94-007-0753-5_1137

Kingdon, D., Serbin, L. A., and Stack, D. M. (2017). Understanding the gender gap in school performance among low-income children: A developmental trajectory analysis. *International Journal of Behavioral Development* 41, 265–274. doi: 10.1177/0165025416631836

Kling, K. C., Hyde, J. S., Showers, C. J., and Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin* 125, 470. doi: 10.1037/0033-2909.125.4.470

Legewie, J., and DiPrete, T. A. (2014). The high school environment and the gender gap in science and engineering. *Sociology of Education* 87, 259–280. doi: 10.1177/0038040714547770

Mac an Ghaill, M. (1994). *The making of men: Masculinities, sexualities and schooling*. Maidenhead: Open University Press.

Marsh, H. W., and Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology* 81, 59–77. doi: 10.1348/000709910X503501

Martino, W. (1999). 'Cool boys', 'party animals', 'squids' and 'poofters': interrogating the dynamics and politics of adolescent masculinities in school. *British Journal of Sociology of Education* 20, 239–263. doi: 10.1080/01425699995434

Morris, E. W. (2008). "Rednecks," "rutters," and 'rithmetic: Social class, masculinity, and schooling in a rural context. *Gender & Society* 22, 728–751. doi: 10.1177/0891243208325163

Morris, E. W. (2011). Bridging the gap: 'Doing gender', 'hegemonic masculinity', and the educational troubles of boys. *Sociology Compass* 5, 92–103. doi: 10.1111/j.1751-9020.2010.00351.x

Muthén, L. K., and Muthén, B. O. (2015). *Mplus* (Version 7) [Computer Software]. Los Angeles, CA: Muthén & Muthén.

Neuburger, S., Jansen, P., Heil, M., and Quaiser-Pohl, C. (2012). A threat in the classroom: Gender stereotype activation and mental-rotation performance in elementary-school children. *Zeitschrift für Psychologie* 220, 61. doi: 10.1027/2151-2604/a000097

O'Mara, A. J., Marsh, H. W., Craven, R. G., and Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist* 41, 181–206. doi: 10.1207/s15326985ep4103_4

Pajares, F., and Valiante, G. (2001). Gender differences in writing motivation and achievement of middle school students: A function of gender orientation? *Contemporary educational psychology* 26, 366–381. doi: 10.1006/ceps.2000.1069

Pascoe, C. J. (2007). *Dude, you're a fag: Masculinity and sexuality in high school*. Berkeley, CA: University of California Press. doi: 10.1525/9780520941045

Quatman, T., and Watson, C. M. (2001). Gender differences in adolescent self-esteem: An exploration of domains. *The Journal of genetic psychology* 162, 93–117. doi: 10.1080/00221320109597883

Reilly, D., Neumann, D. L., and Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist* 74, 445. doi: 10.1037/amp0000356

Rhodes, J., Roffman, J., Reddy, R., Fredriksen, K., and Way, N. (2004). Changes in self-esteem during the middle school years: A latent growth curve study of individual and contextual influences. *Journal of School Psychology* 42, 243–261. doi: 10.1016/j.jsp.2004.04.001

Ridgeway, C. L., and Correll, S. J. (2004). Unpacking the gender system: A theoretical perspective on gender beliefs and social relations. *Gender & society* 18, 510–531. doi: 10.1177/0891243204265269

Roberts, T. A. (1991). Gender and the influence of evaluations on self-assessments in achievement settings. *Psychological bulletin* 109, 297. doi: 10.1037/0033-2909.109.2.297

Ruble, D. N., Greulich, F., Pomerantz, E. M., and Gochberg, B. (1993). The role of gender-related processes in the development of sex differences in self-evaluation and depression. *Journal of Affective Disorders* 29, 97–128. doi: 10.1016/0165-0327(93)90027-H

Rueda-García, N. (2003). *The Challenge of Slums: The Case of Bogotá D.C., Colombia*. Nairobi: United Nations Human Settlements Programme.

Santo, J. B., Bukowski, W. M., Stella-Lopez, L., Carmago, G., Mayman, S. B., and Adams, R. E. (2013). Factors underlying contextual variations in the structure of the self: Differences related to SES, gender, culture, and "majority/nonmajority" status during early adolescence. *Journal of Research on Adolescence* 23, 69–80. doi: 10.1111/j.1532-7795.2012.00793.x

Sebastian, C., Burnett, S., and Blakemore, S. J. (2008). Development of the self-concept during adolescence. *Trends in cognitive sciences* 12, 441–446. doi: 10.1016/j.tics.2008.07.008

Silverman, W. K., La Greca, A. M., and Wasserstein, S. (1995). What do children worry about? Worries and their relation to anxiety. *Child development* 66, 671–686. doi: 10.2307/1131942

Skaalvik, E. M., and Valås, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *The Journal of Experimental Education* 67, 135–149. doi: 10.1080/00220979909598349

Statistics Canada. (2002). Tables 11–10-0012-01. Distribution of total income by census family type and age of older partner, parent or individual. Ottawa, ON: Statistics Canada.

Van Houtte, M. (2004). Why boys achieve less at school than girls: The difference between boys' and girls' academic culture. *Educational Studies* 30, 159–173. doi: 10.1080/0305690320001598 04

Velásquez, A. M., Bukowski, W. M., and Saldarriaga, L. M. (2013). Adjusting for group size effects in peer nomination data. *Social Development.* 22, 845–863. doi: 10.1111/sode.12029

Voyer, D., and Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychological bulletin* 140, 1174. doi: 10.1037/a0036620

World Health Organization [WHO]. (2017). *Gender, equity and human rights*. Available online at: http://www.who.int/gender-equity-rights/understanding/gender-definition/en/ (accessed October 17, 2017)

Wang, J., and Wang, X. (2019). *Structural equation modeling: Applications using Mplus*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9781119422730

Yeager, D. S., and Krosnick, J. A. (2011). Does mentioning "some people" and "other people" in a survey question increase the accuracy of adolescents' self-reports? *Developmental Psychology* 47, 1674–1679. doi: 10.1037/a0025440

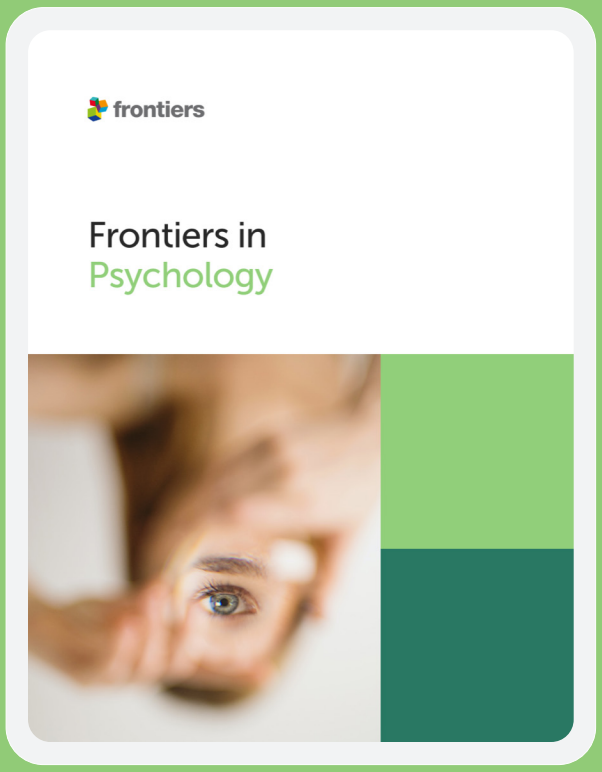
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact