frontiers | Research Topics

Initiatives to raise young people's interest and participation in STEM

Edited by
Milagros Sainz, Katja Upadyaya and Sergi Fàbregues

Published in
Frontiers in Psychology

ISSN 1664-8714
ISBN 978-2-83251-802-1
DOI 10.3389/978-2-83251-802-1

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

Initiatives to raise young people's interest and participation in STEM

Topic editors
Milagros Sainz — Open University of Catalonia, Spain
Katja Upadyaya — University of Helsinki, Finland
Sergi Fàbregues — Open University of Catalonia, Spain

Citation
Sainz, M., Upadyaya, K., Fàbregues, S., eds. (2023). *Initiatives to raise young people's interest and participation in STEM*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83251-802-1

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 16 February 2023
DOI 10.3389/fpsyg.2023.1151715

OPEN ACCESS

EDITED AND REVIEWED BY
Cheng Yong Tan,
The University of Hong Kong,
Hong Kong SAR, China

*CORRESPONDENCE
Milagros Sáinz
✉ msainzi@uoc.edu

SPECIALTY SECTION
This article was submitted to Gender, Sex and Sexualities, a section of the journal Frontiers in Psychology

RECEIVED 26 January 2023
ACCEPTED 31 January 2023
PUBLISHED 16 February 2023

CITATION
Sáinz M, Upadyaya K and Fàbregues S (2023) Editorial: Initiatives to raise young people's interest and participation in STEM. *Front. Psychol.* 14:1151715.
doi: 10.3389/fpsyg.2023.1151715

Editorial: Initiatives to raise young people's interest and participation in STEM

Milagros Sáinz[1]*, Katja Upadyaya[2] and Sergi Fàbregues[3]

[1]Internet Interdisciplinary Institute, Universitat Oberta de Catalunya, Barcelona, Spain, [2]Minds Hub Research Group, Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland, [3]Department of Psychology and Education, Universitat Oberta de Catalunya, Barcelona, Spain

KEYWORDS

interest, participation, gender, Science, Technology, Engineering, and Mathematics (STEM), initiatives

Editorial on the Research Topic
Initiatives to raise young people's interest and participation in STEM

The low share of women in Science, Technology, Engineering, and Mathematics (STEM) is a complex global phenomenon that requires further investigation, since it affects millions of women and girls worldwide. A broad range of interventions, based on a diversity of disciplinary, theoretical, and methodological approaches, have been conducted in different countries and contexts to encourage the participation of young women in various STEM disciplines, especially in those where women remain dramatically underrepresented, such as engineering and physical science. These interventions provide practitioners and policymakers with best practices to tackle the gender gap in STEM careers and professions, promote girls' positive attitudes towards scientific and technological fields, and identify barriers which stand in the way of higher female achievement in science and technology subjects, among other actions. To ensure that these interventions are effective, feasible, and well accepted by their participants, intervention studies grounded in quantitative, qualitative, and mixed methods have been developed.

In the present Research Topic, a variety of aspects related to the under-representation of women in STEM have been addressed with a group of 11 high-quality papers related to gender-based intervention studies. These papers include rigorous empirical studies, methodological papers, and systematic reviews describing initiatives or programs to overcome the gender gap in the STEM educational and career pathways through the following aspects:

- Focusing on research questions and/or objectives related to the gender gap in access and progression of STEM education.
- Drawing on one or more theoretical approaches (i.e., person-environment fit theory, RIASEC model of vocational interests, expectancies and values, social role theory, project-based learning principles, etc.).
- Addressing different stages of educational pathways, including primary, secondary, and higher STEM education.
- Using various methodological approaches to design and evaluate their implementation and effectiveness.
- Discussing the sustainability and long-term effects of the interventions.
- Treating the intersection of gender and other factors, such as areas of study, country of origin, family socioeconomic status, and attained educational level.

The first paper "*Impact of interest congruent on study outcomes*" tackles how social and aspirational congruence interest of a group of German university students are related to students' persistence, performance, and satisfaction in six different study areas, including STEM (Ertl et al.).

In the second paper "*I am done with this. Women dropping out of engineering majors*" the authors conducted a qualitative study with a group of Spanish engineering students, where the main factors (i.e., the influence of stereotypes, lack of role models, excessive academic workload or a hostile class environment) pushing women to drop out of engineering education were identified (González-Pérez et al.).

The third paper entitled "*Girls get Wise. A programming model for engaging girls in STEM*" describes the features and evaluation process of a long-term Canadian university-based program aimed at engaging girls in STEM. Through the use of hands-on interactive STEM activities, this program provides an opportunity for young women to showcase their talents and excitement for science-based topics (Franz-Odendaal and Marchand et al.).

The paper "*On the Design and Validation of Assessing Tools for Measuring the Impact of Programs Promoting STEM Vocations*" addresses the design and validation of an instrument to evaluate how an informal learning initiative developed in Spain promotes Science, Technology, Engineering, and Mathematics STEM vocations among secondary students, their families (parents), and secondary teachers (Herce-Palomares et al.).

The paper "*Perception of work in the IT sector among men and women—A comparison between IT students and IT professionals*" examines gender differences in goal congruence, sense of belonging, and self-efficacy in IT among a group of Polish IT and non-IT workers as well as university students (Pyrkosz-Pacyna et al.).

In the paper "*Interventions to increase young people's interest in STEM. A scoping review*" the authors examine the main characteristics and effectiveness of intervention studies aiming at encouraging secondary school students' interest in STEM over the past 20 years, with a particular focus on female students. Twenty-five studies were also identified as best practices for their design and evaluation characteristics (Sáinz et al.).

The paper "*Associations between adolescent students' multiple domain task value-cost profiles and STEM aspirations*" examines the task value and cost profiles of Finnish middle school students in association with STEM aspirations, and investigates gender differences, using latent transition analysis as a methodological approach (Vinni-Laakso et al.).

The paper "*Gender biases in the training methods of affective computing: Redesign and validation of the Self-Assessment Manikin in measuring emotions via audiovisual clips*" analyzes the development and experimental testing of a graphic design tool for the labeling of emotions free of gender biases (Sainz-de-Baranda Andujar et al.).

The paper "*Gendered difference in motivational profiles, achievement, and STEM aspiration of elementary school students*" uses latent transition analysis to look into gender differences in motivation profiles and their influence on achievement and STEM aspirations over time in a sample of Finnish elementary students (Olive et al.).

The paper "*Intervention initiatives to raise young people's interest and participation in STEM*" examines, using two interventions developed with randomized control trials, how to increase science interest and participation in a group of elementary and secondary school students in the United States (Schneider et al.).

Finally, the paper entitled "*Use of mixed methods research in intervention studies to increase young people's interest in STEM: A systematic methodological review*" examines how the use of a mixed methods approach enhances the comprehensiveness and robustness of an intervention design attempting to raise students' and girls' interest in STEM (Fàbregues et al.).

Author contributions

MS wrote the draft version of the editorial. KU and SF made suggestions. All authors approved the submitted version.

Acknowledgments

We are grateful to all the authors, reviewers, and guest editors for their wonderful insights and collaboration in putting together this Research Topic.

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

ORIGINAL RESEARCH
published: 04 March 2022
doi: 10.3389/fpsyg.2022.816620

Impact of Interest Congruence on Study Outcomes

Bernhard Ertl[1], Florian G. Hartmann[2] and Anja Wunderlich[1]*

[1] *Learning and Teaching With Media, Department of Human Sciences, Universität der Bundeswehr München, Neubiberg, Germany,* [2] *Methodology in the Social Sciences, Department of Human Sciences, Universität der Bundeswehr München, Neubiberg, Germany*

Grounding on Holland's RIASEC model of vocational interests and the respective assumptions on person-environment fit (congruence), this paper focuses on how congruence is related to study outcomes, especially students' persistence, performance, and satisfaction. The paper distinguishes the measure of congruence with respect to social congruence (SOC) (interest fit with the study mates) and aspirational congruence (ASP) (interest fit with the occupation aspired) and also distinguishes the effects of congruence for gender and six different study areas including Science, Technology, Engineering, Mathematics (STEM), medicine, economics, education, and languages. The paper analyses 10,226 university freshmen of the German National Educational Panel Study (NEPS) and follows them longitudinally with respect to their study outcomes. The results show that students' persistence was more related to SOC than to ASP, especially for male students. Furthermore, SOC was particularly important for students in STEM areas. Regarding performance, however, ASP was more important. Here, we notably found correlations for STEM subjects with a balanced proportion of female students. Regarding satisfaction, mainly marginal correlations could be found. The results indicate conceptual differences between social and aspirational congruence as well as specific effects for gender and study area. While research might take this into account by specifically developing their models for different study areas, career counseling may reflect on the different significance of the interest-based person-environment fit for different study areas. Initiatives for raising young people's participation in STEM should therefore specifically focus on students that have high chances to develop interest profiles that are congruent to STEM rather than students who show profiles which already indicate a low congruence.

Keywords: vocational interests, Holland model, university freshmen, congruence, vocational aspiration

OPEN ACCESS

Edited by:
Milagros Sainz,
Open University of Catalonia, Spain

Reviewed by:
Carolina Cárcamo,
National Open and Distance University, Colombia
Katja Upadyaya,
University of Helsinki, Finland

****Correspondence:***
Bernhard Ertl
bernhard.ertl@unibw.de

Specialty section:
This article was submitted to Gender, Sex and Sexualities, a section of the journal Frontiers in Psychology

Received: *16 November 2021*
Accepted: *19 January 2022*
Published: *04 March 2022*

Citation:
Ertl B, Hartmann FG and Wunderlich A (2022) Impact of Interest Congruence on Study Outcomes. Front. Psychol. 13:816620. doi: 10.3389/fpsyg.2022.816620

INTRODUCTION

Choosing an academic major and a corresponding occupation are important decisions determining students' further course of life (Elder, 2002). Person-environment fit (P-E fit) theories suggest that career-related choices are promising if they are based on individual traits (Su et al., 2015) such as vocational interests as defined by Holland (1997). Being one of the most prominent P-E fit approaches (Ott-Holland et al., 2013; Juntunen et al., 2019), Holland's (1997) RIASEC model claims that students who choose an environment that is congruent to their interests should be satisfied, perform well, and persist. Incongruent choices, on the other hand, are supposed to ultimately lead to students dropping out, which causes personal costs such as a lower self-esteem (Hoeschler and Backes-Gellner, 2017) or forgone earnings (Schneider and Yin, 2011) as well as societal costs such as resources invested in vain at institutions of higher education, a lack of skilled professionals, or the loss of tax revenues (Sarcletti and Müller, 2011; Schneider and Yin, 2011; Neugebauer et al., 2019). On an individual level, dropping out can be a sensible decision at some point in time and may result in a degree in another (and better fitting) subject area (Donohue, 2006), but picking an optimal environment right from the start appears to be the better option (Allen and Robbins, 2008). Hence, research looks for preventive measures to avoid dropouts or detours, and interest congruence is considered a decisive predictor.

Nye et al. (2012) carried out a meta-analysis and found a moderate effect of interest congruence on grades and persistence in academic samples. Results of meta-analyses and reviews investigating the effect on satisfaction show rather small correlations especially in academic settings (Assouline and Meir, 1987; Tranberg et al., 1993; Tsabari et al., 2005; Hoff et al., 2020). Recent studies tried to explain such weak relations by moderators, most of which can be assigned to one of three variable types: personal characteristics such as age or gender, characteristics of the environment (such as homogeneity of the work environment regarding values or interests), and methods to conceptualize or measure interest congruence (Fu et al., 2019).

Congruence in Stereotyped Domains

While such studies provide evidence for the effects of congruence in general, much less is known about the effects of congruence for females and males in stereotyped domains like STEM (Science, Technology, Engineering, Mathematics). Gottfredson (2005) introduced the *sextype* of an occupation that may be male, neutral, or female and discussed that individuals may rule out some occupations because of having the wrong sextype. She calls it as a high level of concern if students compromise their interests (respectively their interest congruence) for going into a profession with a better fitting sextype. This may especially apply for the STEM area that has several subjects showing an under-representation of female students (see e.g., Ertl et al., 2017), but also in the education area for male students. Regarding STEM, we focus on the narrow definition also applied in Ertl et al. (2017) that primarily includes the natural sciences, because these are usually considered of having a male sextype. This is different for medicine that is recently more and more seen as a female domain. However, STEM can further be distinguished, analogous to Ertl et al. (2017), in subjects with a low proportion of female students (STEM-L with below 30% of female students, e.g., physics, engineering) for which the sextype *male* corresponds to the proportion of female students and subjects with a medium proportion of students (STEM-M with 30–70% female students) with subjects like mathematics that still are perceived as male although the proportion of male and female students is almost balanced. Distinguishing these areas has some implications e.g., for measures and interventions for promoting females going into STEM. Interest based occupational inventories for career development, e.g., the EXPLORIX (Jörin Fux et al., 2003) or the SDS (Holland et al., 1973), allow to identify female students with good chances for persisting STEM and help thereby to focus measures and interventions on these students that respond best on them. While for students in STEM-M, the regular set of measures may be sufficient because they find plenty of same-sex mates in the courses, this is different for female students in STEM-L that often feel belonging uncertainty (see e.g., Deiglmayr et al., 2019; Höhne and Zander, 2019). Thus, measures for STEM-L may therefore need an additional focus on mentoring (e.g., Kricorian et al., 2020) and networking.

The current study contributes to the investigation of the congruence-outcome relation in academic settings to evaluate how far this measure can inform interventions to select adequate candidates. As academic majors differ with regard to multiple characteristics such as career choice motivations, career choice options, or barriers that have to be overcome (e.g., gender stereotypes), the impact of congruence is supposed to be different for different majors (Tracey et al., 2012; Le et al., 2014; Etzel and Nagy, 2016; Nguyen et al., 2016; Schelfhout et al., 2019b). Therefore, the paper aims at disclosing the effect of interest congruence on study outcomes in different majors including STEM fields with different gender distributions. Building upon recent studies, a holistic investigation is carried out, which includes applying different conceptualizations of congruence and using a sophisticated method to measure congruence and longitudinal large-scale data comprising 14 waves and 8 years.

HOLLAND'S PERSON-ENVIRONMENT FIT THEORY

The RIASEC model (Holland, 1997) describes vocational interests using six different dimensions comprising preferences for specific activities. The *Realistic (R)* dimension includes the preference for practical and technical activities producing concrete results, the *Investigative (I)* dimension comprises of the interest in intellectual and scientific activities studying complex problems, the *Artistic (A)* dimension includes favoring open and unstructured activities demanding creativity and yielding art products or forms, the *Social (S)* dimension consists of social activities like teaching or caring, the *Enterprising (E)* dimension contains the preference for activities that aim at convincing or manipulating other people by verbal or other means, and the *Conventional (C)* dimension conclusively comprises of the interest in regular activities such as the orderly and systematic handling of data. These six dimensions are usually called RIASEC dimensions and are not only used to describe people but also work-related environments such as occupations or majors so that both people and environments can be assigned a RIASEC interest profile. According to Holland's (1997) *calculus hypothesis*, the six dimensions are not independent from each other but can be arranged hexagonally with their spatial distances indicating their psychological similarity. For example, the Social type is more similar to the Enterprising type than to the Conventional type and it is antagonistic to the Realistic type. There is broad evidence for the calculus hypothesis in U.S. samples

(Tracey and Rounds, 1993; Rounds and Tracey, 1996) as well as for other countries such as Germany (Nagy et al., 2010).

The *congruence hypothesis* now states that people who choose an environment of the RIASEC profile which is similar to their personal RIASEC profile show favorable outcomes like satisfaction, performance and persistence. This hypothesis has been comprehensively studied and, overall, has been well confirmed by meta-analyses (e.g., Van Iddekinge et al., 2011; Nye et al., 2012, 2017; Hoff et al., 2020). As mentioned before, there are indications that the congruence-outcome relation is influenced by several moderators. For example, there are manifold methods to conceptualize and to measure congruence affecting the magnitude of the congruence-outcome correlation (e.g., Assouline and Meir, 1987).

According to Muchinsky and Monahan (1987) two *concepts of congruence* can be distinguished. The first, *supplementary fit*, is the similarity between the individual and the people in the environment. The second, *complementary fit*, focuses on the resources or demands of an environment and is given if a person's abilities meet these demands or if a person's needs are met by the supplied resources (Kristof-Brown et al., 2005; Su et al., 2015). Scholars disagree on whether interest congruence is more of supplementary or complementary nature. For example, while Henry (1989) states that "Holland's theory is the prototypical supplementary model in which individuals pursue careers which supplement their interests" (p. 38), there are indications that interest congruence follows a rather complementary fit approach (Wiegand et al., 2021).

Regarding the *measurement of fit*, there is a distinction of *direct and indirect measures* (Kristof-Brown et al., 2005). A direct measure would be the perceived fit of an individual. Indirect measures require the assessment of both the person's and the environment's characteristics. In addition, indirect measures can be *subjective*, when the characterization of both the person and the environment are based on self-reports by the same person, or *objective*, when the description of the person and the environment are based on different sources. With respect to objective measures, the assessment of the environment can be based on the aggregated interests of the people in the environment (*incumbents*) or on the judgments by *experts* (Rounds et al., 1999). Once the data of the person and of the environment are available, no matter whether the subjective or objective approach was followed, congruence in the context of Holland's (1997) theory can be calculated alternatively with more recent measures being superior. Here, older *measures* of congruence use only three or less RIASEC dimensions that are dominant for a person or an environment, respectively (Camp and Chartrand, 1992; Brown and Gore, 1994). These measurement approaches have been criticized since the full RIASEC interest profile information is not used and potential outcome-relevant information is ignored (Dik et al., 2010; Tracey and Sodano, 2013). Measures that are recently used take all six RIASEC dimensions into account with the *Euclidean distance* being especially theoretically grounded since it also counts in the aforementioned calculus hypothesis. Following the assumption that a RIASEC profile can be represented by a vector in a two-dimensional space (Prediger, 1982; Eder, 1998), the Euclidean distance measures the distance between the personal RIASEC vector and the environmental RIASEC vector so that smaller values indicate higher congruence. In the current study, the Euclidean distance is used to measure congruence.

META-ANALYTIC FINDINGS ON THE EFFECT OF INTEREST CONGRUENCE ON ACADEMIC OUTCOMES

Satisfaction, performance, and persistence are central education- and work-related outcomes whose relation to vocational interests has been widely studied. Some of the reviews and meta-analysis summarizing this research focused solely on the occupational context (Van Iddekinge et al., 2011; Hoff et al., 2020) while others also carried out analyses for academic environments. This has been done with regard to the association of interest congruence with satisfaction (Assouline and Meir, 1987; Tranberg et al., 1993; Tsabari et al., 2005) as well as with performance and persistence (Nye et al., 2012).

Performance and Persistence

Considering academic samples, Nye et al. (2012) summarize over 60 years of research on the effects of vocational interests on performance and persistence. Overall, they found that interests correlated moderately with performance and persistence for employed samples as well as for academic samples (baseline corrected correlations of interest congruence with performance and persistence of 0.30 and 0.36 in employed samples and 0.30 and 0.34 in academic samples). For the current investigation, we took a closer look at those studies that analyzed interest congruence in academic samples and were explicitly based on Holland's (1997) theory. Here, we focused on three variables that could moderate the congruence-outcome relation, namely methods to measure congruence, gender (as a personal characteristic), and diversity of majors (as a characteristic of the environment) (see Fu et al., 2019).

With regard to the measurement of congruence, the studies summarized by Nye et al. (2012) predominantly used dated methods. For example, Chartrand et al. (1992) used the M Index (Iachan, 1984), which considers only the three dominant RIASEC dimensions of the person and the three dominant dimensions of the environment. Other studies even used only one dimension of the person and one dimension of the environment and ignored the remaining RIASEC dimensions (Bruch and Krieshok, 1981; Henry, 1989). Since such studies do not use full interest profile information to depict the interest structure of the person and the environment, they may be a suboptimal database to estimate the effects of interest congruence. They are furthermore able to provide only ordinal congruence estimations (see Hartmann et al., 2021). As a consequence, the congruence-outcome relation could be different when sophisticated methods are used. Based on the meta-analysis by Nye et al. (2012), this question remains unanswered.

With respect to gender, some studies investigated only males (Bruch and Krieshok, 1981) or only females (Camp and Chartrand, 1992). Studies examining both male and female students mostly found similar effects (Henry, 1989; Leuwerke et al., 2004). However, the study by Nichols and Holland (1963) found interest-outcome relations varying with gender. For example, the relation between the preference for realistic occupations and literary achievement was –0.37 for males and –0.08 for females while the relation between the preference for realistic occupations and (rare) scientific achievement was –0.06 for males and 0.43 for females.

Focusing on the diversity of majors in the samples, part of the studies investigated students from only one single academic major and came to different conclusions about the congruence-outcome relation. For example, Chartrand et al. (1992) only analyzed the vocational interests of psychology students and found no relation between interest congruence and students' GPA in psychology. Bruch and Krieshok (1981) studied students with engineering majors especially requiring Investigative interests. They only considered students who showed high Realistic or high Investigative interests. In line with the congruence hypothesis, students with high Investigative interests showed higher persistence and also attained better grades than students showing high Realistic interests. These studies focusing on different, and in each case, homogenous groups of students indicate that the academic major might be a moderator of the congruence-outcome relation. Studies considering multiple academic majors (e.g., Schmitt et al., 2008) yielded significantly lower effects compared to studies focusing a single major (Nye et al., 2012).

Overall, the meta-analysis by Nye et al. (2012) indicates that interest congruence is an important predictor of performance and persistence in the academic context and can be a preventive factor with regard to dropouts or detours. However, many studies summarized, used dated and rather coarse methods to measure congruence, and it remains unclear whether the effects are equal for male and female students in different academic majors.

Satisfaction

With regard to the relation between interest congruence and satisfaction, a few reviews and meta-analysis have been carried out (Assouline and Meir, 1987; Tranberg et al., 1993; Tsabari et al., 2005). These investigations found no significant correlations within the occupational context and negligible or non-significant correlations between interest congruence and satisfaction in the academic context. In detail, Assouline and Meir (1987) found a mean correlation of 0.098 based on six correlations, Tranberg et al. (1993) found a mean correlation of 0.095 based on five correlations, and Tsabari et al. (2005) only cited one study analyzing academic satisfaction, which showed a correlation of –0.033. Recently, Hoff et al. (2020) criticized these meta-analyses and carried out another meta-study overcoming some weaknesses of the criticized ones. For example, they analyzed far more studies resulting in sufficient statistical power and revealing a statistically significant correlation between interest fit and job satisfaction of 0.19 (95% CI: 0.16,0.21). Unfortunately, they did not consider studies investigating academic samples.

To sum up, according to previous meta-analyses, evidence for the congruence-satisfaction relation is weak; however, this might be due to methodological issues or other moderating variables such as gender or major. The results regarding performance, persistence, and satisfaction also indicate that interest congruence may have effects of different strengths on different academic outcomes, namely a stronger effect on performance and persistence than on satisfaction.

RECENT EVIDENCE

With few exceptions, current studies use sophisticated methods to measure congruence such as profile correlation or the Euclidean distance, which take all six RIASEC dimensions into account (e.g., Tracey et al., 2012; Nguyen et al., 2016; Kim and Beier, 2020). These studies found interest congruence to be related to performance (Tracey et al., 2012; Nye et al., 2018), major persistence (Allen and Robbins, 2008; Tracey et al., 2012; Le et al., 2014; Le and Robbins, 2016; Nguyen et al., 2016; Kim and Beier, 2020), and major satisfaction (Bai and Liao, 2019). Studies that took gender as a moderator into account indicated that the effects of interest congruence on persistence (in STEM) is similar for men and women (Le et al., 2014; Le and Robbins, 2016). As with studies considering gender as a moderator, studies considering academic majors as a moderator are rare. Tracey et al. (2012) focused only on a specific aspect of academic majors, namely environmental constraint (i.e., interest homogeneity within an academic major), and concluded that the relation between interest congruence and academic outcomes is stronger for majors that are more constraint, i.e., show more homogeneity regarding the RIASEC profiles of the people in the majors (incumbents). Le et al. (2014) carried out separate analyses for two groups of STEM majors: STEM Science and STEM Quantitative (technology, engineering, and mathematics). Although they found similar main effects of interest congruence on persistence, effects regarding the interaction between abilities and interest fit were different. Nguyen et al. (2016) carried out separate analyses for students of biology and students of chemistry. Using the Euclidean distance as a congruence measure, they only found interest congruence to be related to persistence for chemistry but not for biology. Based on their results Nguyen et al. (2016) concluded that "the factors that impact university success and retention vary across majors, complicating attempts to address retention efforts at universities; a broad, discipline-unspecific approach will not suffice" (p. 12). In addition, it is possible that two moderating variables, such as gender and major, interacted with one another. For example, Le et al. (2014) referred to Heilman et al. (2004) and argued that women in STEM fields have to overcome barriers such as gender stereotypes including negative reactions in case of success regarding tasks that are male gender-typed. These preconditions (or external factors) could weaken the role of interest congruence and its impact on academic outcomes for female students in such majors (see also Lent, 2013; Fu et al., 2019).

THE CURRENT STUDY

To sum up, recent research gives some indications that the effects of congruence on academic outcomes vary with respect to gender and subject area although systematic comparisons are missing so far. To shed light on this area, this study will analyze the effects of congruence with a specific focus on gender and subject area. Our main focus lies on the STEM areas with a low and medium proportion of female students, but we will contextualize these results with further areas for being able to discuss whether the effects revealed are STEM specific or rather general. Therefore, the present study grounds its analyses on a national large-scale study that enables the analysis of effects in general as well as specifically with respect to gender and different academic areas. The present study includes two conceptualizations of congruence: social congruence (SOC) that depicts the congruence between the interests of the individual and the interests of the respective mates in the study subject (incumbents) and aspirational congruence (ASP) that depicts the congruence between the individual's interests and the professions aspired. For both conceptualizations, a sophisticated method is used to quantify the extent of congruence by calculating the Euclidean distance, which is grounded on Holland's (1997) RIASEC model and considers full profile information. This also allows a better estimate of mean differences, e.g., between persisters and non-persisters rather than only rank-order correlations. The study furthermore covers several outcome measures like completing the studies, the respective grades, and study satisfaction.

RESEARCH QUESTIONS

This study analyses the effect of interest congruence (at study entry) on study outcomes. The approach thereby is to focus first on the whole cohort, then specifically on male and female students, and then on differences between the study areas.

The first research question focuses students' *persistence*, i.e., if students finished their degree program successfully, if they failed, or if they left their studies with a different state. As these states of persistence are categorical variables, RQ 1 analyzes to what extent students of these persistence state groups distinguish with respect to their congruence. The respective research question is:

RQ1: To what extent do students with different states of persistence distinguish with respect to social and aspirational congruence?

> RQ1a: To what extent do male and female students with different states of persistence distinguish with respect to social and aspirational congruence?
> RQ1b: To what extent do students of different subject areas with different states of persistence distinguish with respect to social and aspirational congruence?

> *Hypothesis 1*: Previous research, summarized by the meta-study of Nye et al. (2012) provided evidence for correlations between congruence and persistence. Therefore, we hypothesize successful/persisting students being more congruent than less successful. Although literature indicates effects for gender (Le et al., 2014) and study subject (Nguyen et al., 2016), the body of previous research is not strong enough to deduct clear hypotheses regarding either of both aspects.

Besides persistence, also performance will be considered:

RQ2: To what extent do social and aspirational congruence correlate with study grade (*performance*)?

> RQ2a: To what extent do social and aspirational congruence correlate with study grade for male and female students?
> RQ2b: To what extent do social and aspirational congruence correlate with study grade for the different study areas?

> *Hypothesis 2*: According to Nye et al. (2012), we expect that congruence correlates positively with study grade (performance).

The third research question will focus on satisfaction:

RQ3: To what extent is students' satisfaction related to social and aspirational congruence?

We will investigate this research question for male and female students (RQ3a) as well as for students from different subject areas (RQ3b). With respect to satisfaction, we would expect positive correlations in accordance with Holland's (1997) congruence hypothesis which means that a better congruence is correlated with a higher satisfaction. However, previous research showed ambiguous results in this regard. Therefore, we avoided stating a clear hypothesis and instead assumed that study subjects and gender may be a key to interpret the ambivalent results.

MATERIALS AND METHODS

The data set for this study comes from the German National Educational Panel Study (NEPS; Blossfeld et al., 2011; see also acknowledgments) and its starting cohort 5 that focuses on first year students (SC5:14:0.0). This cohort started their studies in autumn 2010 (FDZ-LIfBi, 2021b), and is still actively being followed. Students entered the survey with wave 1 in autumn 2010 and were then surveyed in each wave either per computer assisted telephone interview (CATI), computer assisted web interview (CAWI), or, e.g., for competence tests, sometimes a part of the students was tested in presence. An overview on the survey waves and times of survey can be found at FDZ-LIfBi (2021b). Some student information does not fit into the wave format and is represented in an episode format in the dataset, e.g., participation in study programs, internships, or work contracts. These were coded as episodes with a starting point and an end date, the study program in which a student participated and a status of termination of a study program (see FDZ-LIfBi, 2021a,b). Thus, if a student was enrolled in two study programs concurrently then the episode data would contain two episodes for these students with overlapping dates. If students mentioned that they dropped out, they were given a specific reason for dropout questionnaire[1]. However, the data set contains a notable number of students

[1]More information on students' reasons for dropout could be found at Mouton et al. (2020).

with open episodes; for them it remains unclear if they dropped out, if they finished their studies, or if they were unavailable for other reasons. We will tackle this phenomenon by focusing on two different measures for persistence: (a) we first will analyze whether students finished their initial study episode successfully or whether they failed and (b) then we will check how far students finished any study episode successfully or whether they mentioned that they dropped out.

The current data set includes 14 survey waves, and wave 14 was surveyed in autumn 2018. The study analyzed several variables that were surveyed in specific waves, especially interests (wave 1), aspirations (wave 1), and satisfaction (wave 3 that was collected in spring, 2012 after about 18 months of study, and wave 5 that was collected one year later in spring 2013; FDZ-LIfBi, 2021b). The dataset also included event specific variables like the completing or failing of a study episode with its respective final grade. This longitudinal perspective was essential for being able to reveal long time effects but suffers from issues like panel attrition, which resulted in far less students providing information in wave 14.

Sample

Within this NEPS cohort of first year students, the study specifically focuses on students with the birth years from 1988 to 1991 to have a homogeneous age range that includes the majority of first year students (79%). Consequently, they were aged between 18 and 22 at study entry and respectively between 21 and 25 at wave 5. Within this age range, the study further narrows its focus on six major study areas: STEM with a low proportion of female students (STEM-L; less than 30% female students; 2,979 students; e.g., physics, engineering, computer science; see **Supplementary Material 1**), STEM with a balanced proportion of male and female students (STEM-M; between 30 and 70% female students; 2,457 students; e.g., mathematics, biology, chemistry), medicine (498 students; balanced; mainly general medicine and dentistry), economics (1,369 students; balanced; only the group of economics), education with a high proportion of female students (723 students; more than 70% female students; including education and welfare), and languages (2,200 students; high; German, English, and Roman language). In total, the study includes 10,226 students, 4,269 males and 5,957 females which are 57% of the initial sample.

Congruence Measures

Both congruence measures build on an assessment of vocational interests. These were surveyed at wave 1 by the Interest Inventory Life Span (IILS-II scales; von Maurice and Nagy, 2009; Wohlkinger et al., 2011; see also FDZ-LIfBi, 2021a, pp. 699–704). This inventory covered each dimension with three items in a Likert scale format. Internal consistencies varied from $\alpha = 0.523$ to $\alpha = 0.749$ (Cronbach's α for Realistic: $\alpha = 0.704$; Investigative: $\alpha = 0.625$; Artistic: $\alpha = 0.629$; Social: $\alpha = 0.749$; Enterprising: $\alpha = 0.523$; Conventional: $\alpha = 0.561$). Considering that these α relate to short scales for a large-scale panel study, they can be estimated as acceptable for analysis (see e.g., Rammstedt and Beierlein, 2014; Ziegler et al., 2014; Ertl et al., 2020). By applying the Randall program (see Tracey, 1997), we confirmed the hexagonal structure of the dimensions (see Hubert and Arabie, 1987; CI = 0.81; $p = 0.017$; see also Ertl and Hartmann, 2019). Based on the hexagonal structure of the interests, an interest vector comprising all six dimensions was created for each participant according to the suggestions of Prediger (1982) and Eder (1998).

Regarding SOC, a second interest vector for the peer group in a specific subject (e.g., civil engineering, German language) was created. Therefore, interests of all students within a subject were aggregated and the vector was built similarly as for the individual but based on the mean level of the interests of the respective peer group in a specific subject.

For ASP, a third vector was built based on the interest profiles of the occupations listed in the O*Net (2018) database. O*Net contains expert ratings for each profession according to the six interest dimensions. Based on the occupations that students aspired in wave 1, the respective interest vector was assigned to each student. For that, students' aspirations were first classified according to ISCO-08 (International Labour Office, 2020) classification (e.g., the code 2111 was assigned for *Physicists and Astronomers*). Then, the respective vector from the O*Net data was matched based on the procedure described by Ertl and Hartmann (2019).

Both congruence measures, the SOC as well as the ASP were calculated as the Euclidean distance between the individual and the peer group/aspiration vector as a measure for congruence according to Tracey and Sodano (2013). Notably, low values of this congruence measure indicate a high congruence.

Persistence and Performance

The NEPS dataset provides a table listing all study episodes of a student (FDZ-LIfBi, 2021b). In this, each study episode was documented, e.g., whether a student started with a BA in computing and then continued with a MA, then the start and end dates of both study programs together with the termination status and the respective grade were documented as two study episodes. If this student concurrently enrolled for biology also this was documented as a further episode. This episode data was applied to extract information about students' study outcomes. Regarding this, we distinguish two categories: the outcomes of the initial study episode starting at study entry (Initial episode) and, more general, a students' outcomes within the timeframe of the NEPS panel. We now first look at the initial study episodes starting at study entry. These were analyzed regarding whether students failed (2,143 students) or completed successfully (4,398 students). Moreover, a minor number explicitly mentioned that they didn't finish these study episodes, or it occurred that they ended this study episode within the first term. Because of the low number of cases, we didn't include these outcomes in the analyses. Furthermore, the study episodes were still open for about one third of the sample, e.g., due to panel attrition (see **Supplementary Material 2**).

If students mentioned that they finished a study episode successfully and provided a grade, these grades were taken as performance measure. To ensure comparability, the grades were z-standardized within the respective study subject group. This means that we built a comparison group of e.g., students who were successfully finishing a BA in engineering and for this group we built z-values indicating how many standard deviations a student's grade deviated from the group mean. Of note, lower values mean better grades in the German grading system.

In a broader context regarding the general outcomes, we observed that 1,485 of the 4,398 students completed their initial degree and furthermore earned a follow-up degree, e.g., a Master degree after a Bachelor (see **Supplementary Material 3**). 2,913 students only completed an initial degree. Moreover, we observed that 1,246 students completed another degree but not their initial degree (successful changers). Six hundred and six students filled a reason for dropout questionnaire without earning any degree. Five students previously earned a university degree but not within the NEPS runtime; they will be dropped for further analyses. In conclusion, we don't have any information about a degree or a dropout for almost 40% of the sample. This number seems reasonable when considering effects of panel attrition and the generally quite high dropout numbers in Germany (see Heublein, 2014).

Satisfaction

Students' satisfaction was measured at wave 3 and wave 5 by the scales of Westermann et al. (1996) with the subscales of general study satisfaction (e.g., "I enjoy my degree course"; Cronbach's α for wave 3: α = 0.880; wave 5: α = 0.890), exhaustion (e.g., "Degree course and other obligations are hard to match"; Cronbach's α for wave 3: α = 0.776; wave 5: α = 0.773), and satisfaction with study conditions (e.g., "Wishing better study conditions"; Cronbach's α for wave 3: α = 0.754; wave 5: α = 0.756).

RESULTS

All analyses were calculated with SPSS 25.0 at the remote NEPS site in Bamberg. Missing values were excluded pairwise for each analysis and therefore Ns are provided for each analysis. Confidence intervals for the correlations were calculated by MPlus 8.2 at the remote NEPS site in Bamberg.

Congruence Differences for Students With Different Levels of Persistence

The first research question asked how far students with different levels of persistence distinguish regarding their congruence. This endeavor, however, has several framing conditions as all longitudinal studies suffer from panel attrition, implying that members drop out the panel without further information. It may be that they were no more willing to participate but also that they changed their contact information. These students may have abandoned their studies, but they may have also finished their studies and have just left the panel. Therefore, we will consider this group as unknown missing comprising of a part of students missing at random and a part of students missing systematically. We will report their values but exclude this group in the discussion of significant differences. This is different to explicit study dropout when students ended their studies without graduating and gave information about why abandoning their studies.

Regarding the initial study programs (ISP) that started at wave 1, we analyzed mean differences in the congruence levels for students within the different outcome groups. Here we can see that the students who failed showed a significant worse SOC [$F_{\text{Welch}(1, 3846.545)} = 44.951$; $p < 0.001$; Cohen's $d = 0.18$] and ASP [$F_{\text{Welch}(1, 3094.800)} = 11.504$; $p = 0.001$; Cohen's $d = 0.11$] than their mates that finished successfully. Regarding the general outcomes, we can see significant differences for SOC [$F_{\text{Welch}(4, 2785.974)} = 8.999$; $p < 0.001$] as well as for ASP [$F_{\text{Welch}(4, 2067.079)} = 9.597$; $p < 0.001$] with respect to the different outcome categories (see **Table 1**). *Post Hoc* Tests with Tamhane adjustment revealed that students who were completing their initial degree with or without finishing a follow-up degree showed a significant higher SOC than students that finished after changing their subject or that explicitly dropped out (Cohen's $d_{\text{max}} = 0.24$). With respect to ASP, students who only completed an initial degree showed significantly higher congruence than students who dropped out as well as students that earned a follow up-degree (Cohen's $d_{max} = 0.25$).

When investigating gender specifically (RQ1a), the results of the ISP that started at wave 1 showed a lower SOC for male [$F_{\text{Welch}(1, 1774.717)} = 19.333$; $p < 0.001$; Cohen's $d = 0.18$] and female students [$F_{\text{Welch}(1, 2083.846)} = 25.664$; $p < 0.001$; Cohen's $d = 0.19$] who failed. For ASP, we did not observe an effect for male students ($p = 0.227$) but for female students [$F_{\text{Welch}(1, 1821.894)} = 9.684$; $p < 0.01$; Cohen's $d = 0.12$]. Regarding general outcomes, the significant differences disappear for male students ($p = 0.117/0.141$) while they hold for female students for SOC [$F_{\text{Welch}(4, 1476.769)} = 7.105$; $p < 0.001$] as well as ASP [$F_{\text{Welch}(4, 1159.132)} = 4.634$; $p = 0.001$]. Again, female students who explicitly dropped out showed a significant worse SOC than female students that completed their initial degree program with or without follow-up (Cohen's $d_{max} = 0.32$). With respect to ASP, we again could only see that female students that completed their initial degree showed a better congruence than female students that dropped out (Cohen's $d_{max} = 0.27$).

Regarding RQ1b that distinguishes the different persistence levels with respect to gender and subject area, we would get a quite extensive table. Therefore, we keep all the congruence values including confidence intervals in the **Supplementary Material 4** and just present the *p*-values in **Table 2**. Besides indicating areas where significant differences occur, these *p*-values give also insights in the areas that are far away from significant differences, which means that the not-significance may be not caused by a decreasing sample size but rather by a non-existence of an effect in the context of a certain area.

Looking now more detailed into **Table 2**, we see the differences disappearing for about the half of the subject areas. The *p*-values shown in **Table 2** indicate furthermore that most of the missing significances can be explained by non-existent effects rather than by a shrinking sample size. This applies especially for the areas of medicine, economics, and education. When first examining the outcomes of the initial studies, we can see that students in both STEM areas, males as well as females, are highly reliant on SOC with students passing having a significantly higher congruence than students failing (STEM-L: Total Cohen's $d = 0.22$; male $d = 0.17$; female $d = 0.37$; STEM-M: Total Cohen's $d = 0.23$; male $d = 0.25$; female $d = 0.21$). For the ASP, there was only one significant effect for the languages (Cohen's $d = 0.15$), and this is also primarily going back to the female students (Cohen's $d = 0.19$).

TABLE 1 | Means for social congruence (SOC) and aspirational congruence (ASP) for all students (Total) and separation of male and female students with respect to their initial study program (ISP) and general study outcomes.

Congruence	Total		Male students		Female Students	
ISP outcome	**SOC**	**ASP**	**SOC**	**ASP**	**SOC**	**ASP**
ISP failed	0.372	0.846	0.370	0.920	0.373	0.797
ISP success	0.338	0.812	0.337	0.901	0.339	0.758
Gen. outcome						
Open group	0.344	0.818	0.345	0.905	0.344	0.762
Explicit dropout	0.381	0.873	0.365	0.924	0.396	0.830
Successful changers	0.359	0.826	0.353	0.908	0.362	0.782
Initial degree only	0.339	0.793	0.337	0.884	0.340	0.748
Initial degree + follow up	0.338	0.854	0.337	0.930	0.339	0.789

*Lower values indicate higher congruence. $N_{ISPfailed/SOC}$ = 2,139; $N_{ISPfailed/ASP}$ = 1,646; $N_{ISPsuccess/SOC}$ = 4,393; $N_{ISPsuccess/ASP}$ = 3,303; $N_{OpenGroup/SOC}$ = 3,969; $N_{OpenGroup/ASP}$ = 3,051; $N_{Explicitdropout/SOC}$ = 605; $N_{Explicitdropout/ASP}$ = 448; $N_{Successfulchangers/SOC}$ = 1,243; $N_{Successfulchangers/ASP}$ = 997; $N_{Initialdegreeonly/SOC}$ = 2,908; $N_{Initialdegreeonly/ASP}$ = 2,299; $N_{Initialdegree+followup/SOC}$ = 1,485; $N_{Initialdegree+followup/ASP}$ = 1,004. The group of ISP success is separated into the groups of Initial degree only and Initial degree + follow up in the lower part of the table. See **Supplementary Materials 2, 3** for gender distributions within groups and confidence intervals.*

TABLE 2 | *p*-values for differences with respect to social congruence (SOC) and aspirational congruence (ASP) for all students (Total) and separate for male and female students for the different subject areas with respect to their initial study program (ISP) and general study outcomes.

p-values	Total		Male students		Female Students	
ISP outcome	**SOC**	**ASP**	**SOC**	**ASP**	**SOC**	**ASP**
STEM-L	0.000***	0.831	0.003**	0.473	0.001**	0.414
STEM-M	0.000***	0.118	0.003**	0.205	0.001**	0.672
Med	0.965	0.344	0.407	0.998	0.644	0.360
Eco	0.053	0.841	0.296	0.550	0.096	0.854
Edu	0.432	0.900	0.386	0.769	0.551	0.826
Lang	0.170	0.011*	0.697	0.416	0.218	0.005**
Gen. outcome						
STEM-L	0.101	0.350	0.475	0.076	0.192	0.017*
STEM-M	0.003**	0.263	0.070	0.000***	0.025*	0.308
Med	0.808	0.350	0.750	0.797	0.677	0.562
Eco	0.631	0.551	0.407	0.606	0.152	0.809
Edu	0.794	0.917	0.521	0.955	0.859	0.889
Lang	0.016*	0.230	0.826	0.292	0.023*	0.057

*Because several sub-samples show heterogeneous variances, we will generally report p-values of differences for a robust test of equality means (Welch). Means, SDs, confidence intervals and Ns for the respective subgroups could be found in **Supplementary Material 4**.*
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Regarding general outcomes, we still have the effect for STEM-M indicating that dropout students show lower SOC than students that finish their initial degree with follow-up (Cohen's d = 0.30). When separating this for male and female students, we can observe that the significance levels shrink in a way that for male students the differences are only on a tendency level. For female students we had the phenomenon that, although the overall ANOVA indicates significant differences, the *Post Hoc* was not able to assign these to specific group differences. Of note, male students in STEM-M that completed their initial degree program and a follow up showed better ASP than students dropping out (Cohen's d = 0.47). For the languages, students with explicit dropout display lower SOC than students that complete their initial degree and a follow up study (Cohen's d = 0.39). This, however, is particularly demonstrated by the female students (Cohen's d = 0.40). Finally, female students in STEM-L that proceed into a follow-up study exhibit higher congruence than those who do not continue their education (Cohen's d = 0.54).

Correlations of Congruence and Performance

Looking now at performance (RQ2), we analyze the correlations between congruence and the performance measure study grade. Regarding this we can see that SOC was not related to this measure, neither for the total sample nor for males or females separately (see **Table 3**). ASP showed only marginal, albeit significant, correlations with the final grades – for the whole sample as well as for male and for female students (RQ2a). Notably, lower values mean better outcomes for both congruence and grades. Therefore, positive correlations indicate that a better fit goes along with a better grade. Looking at the confidence

TABLE 3 | Correlations of social congruence (SOC) and aspirational congruence (ASP) with grades for all students (Total) and separation of male and female students.

	Total		Male students		Female Students	
	SOC	**ASP**	**SOC**	**ASP**	**SOC**	**ASP**
Final grade (*r*)	0.025	0.061**	0.011	0.071*	0.035	0.056*
N	4220	3159	1745	1182	2475	1977

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

TABLE 4 | Correlations of social congruence (SOC) and aspirational congruence (ASP) with grades for all students (Total) and separate for male and female students for the different subject areas.

Correlations	Total		Male students		Female students	
Final grade	**SOC**	**ASP**	**SOC**	**ASP**	**SOC**	**ASP**
STEM-L	0.005	0.036	0.019	0.012	–0.059	0.136
STEM-M	0.002	0.127**	–0.010	0.161*	0.008	0.103*
Med	–0.019	0.100	–0.176	0.166	0.053	0.071
Eco	0.064	0.033	0.032	0.073	0.083	0.003
Edu	0.064	0.120*	0.211	0.209	0.054	0.114
Lang	0.049	0.079*	0.033	0.215*	0.054	0.070

Ns and confidence intervals can be found in ***Supplementary Material 6****.*
*$p < 0.05$; **$p < 0.01$.

intervals in **Supplementary Material 6**, we cannot confirm significant differences between male and female students.

When examining different subject areas (RQ2b), we can observe that some correlations were not significant (see **Table 4**). The remaining, however, increased their sizes. Thus, we could observe a correlation pattern for the final grades indicating that the relation between congruence and performance depended on the subject area and gender. ASP showed stable correlations with the grades for STEM-M for the whole group as well as for male and female students. Regarding languages, the total effect was marginal while male students showed a notable effect, and for education, there was only an effect for the total sample rather than for male and female students in detail. Several other correlations showed values greater than 0.1 but were not significant. Thus, we did not interpret them. With regard to SOC, there were no significant effects at all. Overall, the correlation patterns that emerged when different subject areas are distinguished are quite diverse and simultaneously appear to be specific for some measures (see **Supplementary Material 6**).

Correlations of Congruence and Satisfaction

When examining the satisfaction variables (RQ3), we analyzed correlations between congruence and satisfaction. For this RQ, there were some significant but altogether marginal correlations[2]. Study satisfaction in general did not correlate with any mode of congruence, neither at wave 3 nor at wave 5 (see **Table 5**). An exception could be male students at wave 3 where a better congruence went along with a higher study satisfaction (RQ3a). However, students with a lower congruence felt more exhausted — only the SOC of male students at wave 3 did not correlate. Effects regarding study conditions were quite specific. We generally found that a higher SOC coincided with more satisfaction about the study conditions. This was only applicable for the total sample in wave 3 as well as for the total sample and the male and female sub-samples in wave 5. Regarding ASP, the correlation was the other way around: a higher ASP went along with a lower satisfaction with the study conditions for the total sample and the female sub-sample at wave 3 and wave 5. We could only observe that a better ASP went along with a higher satisfaction with the study conditions for male students in wave 5. Notably, all the correlations, although some were highly significant, were below the Cohen (1988) threshold for small effects. The confidence intervals in **Supplementary Material 7**, however, indicate significant differences in the correlations between ASP and study conditions between male and female students at wave 3 as well as at wave 5: While a higher congruence was correlated with a higher study satisfaction for male students, this was opposite and significantly different for female students.

Regarding the subject areas (RQ3b), many of the correlations disappeared (**Table 6** summarizes areas with significant differences). On the significance level, we can observe that most of the significant correlations between either mode of congruence and the sub-scales of satisfaction could be found in both STEM domains. Yet, when focusing on the correlation sizes, we noticed that all of them were below Cohen's (1988) threshold of 0.1 – especially for wave 3. Regarding the STEM subjects, this also applies to wave 5. In wave 5, however, we found a small correlation between SOC and exhaustion for medicine ($r = 0.111$; see **Supplementary Material 5**) and for economics ($r = 0.105$) between ASP and this dimension – both indicating that students with a better congruence are less exhausted. Thus,

[2]When interpreting these effect sizes, we followed Cohen's (1988) well-known suggestion that effects smaller than 0.1 are rather negligible.

TABLE 5 | Correlations between social congruence (SOC) and aspirational congruence (ASP) and satisfaction measures for all students (Total) and separate for male and female students.

Correlations	Total		Male students		Female students	
W3	**SOC**	**ASP**	**SOC**	**ASP**	**SOC**	**ASP**
Study satisfaction	–0.008	–0.008	–0.025	–0.043*	0.003	0.007
Exhausted	0.041***	0.041**	0.022	0.062**	0.054***	0.048**
Study conditions	–0.026*	0.053***	–0.029	–0.034	–0.024	0.050**
W5						
Study satisfaction	–0.018	–0.005	–0.015	–0.040	–0.021	0.013
Exhausted	0.067***	0.049***	0.048**	0.078***	0.080***	0.061***
Study conditions	–0.048***	0.053**	–0.042*	–0.047*	–0.056***	0.054**

Lower values indicate higher congruence.
**p < 0.05; **p < 0.01; ***p < 0.001.*

TABLE 6 | Areas with significant differences for the correlations between social congruence (SOC) and aspirational congruence (ASP) and satisfaction measures for all students and separate for the different study areas.

W3	STEM-L	STEM-M	Med	Eco	Edu	Lang
Study satisfaction	–/–	–/*	–/–	–/–	*/–	–/–
Exhausted	–/*	–/–	–/–	–/–	–/–	–/–
Study conditions	*/**	–/–	–/–	–/–	–/–	*/–
W5						
Study satisfaction	–/–	–/–	–/–	–/–	–/–	–/–
Exhausted	**/–	**/–	*/–	–/*	–/–	–/–
Study conditions	*/**	**/*	–/–	*/–	–/–	–/–

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

although most significances could be found in the area of STEM, only exhaustion at wave 5 showed small effects with congruence for medicine (social) and economics (aspirational).

Separating these down into gender, we can detect that there were several significant correlations for male students for STEM-L – although all below Cohen's (1988) threshold of 0.1. Yet, there was one small significant effect for female students: in wave 3, a better ASP was correlated with more satisfaction with the study conditions in wave 5 (r = –0.109). For STEM-M there were significant correlations only for female students; however, only one notable effect indicated that a higher congruence went along with more satisfaction with the study conditions (r = –0.147; see **Supplementary Material 5**). For medicine, there were no significant correlations except for the female students in wave 5 indicating that students with a higher SOC felt less exhausted (r = 0.137). Male students of economics with a better congruence also felt a higher study satisfaction at wave 3 (r = –0.172) as well as less exhaustion at wave 5 (r = 0.165). Their female classmates showed two significant correlations, however, with marginal effect sizes. In education, no significant effects were found and in the languages, there was only one with a marginal effect size.

DISCUSSION

In line with Holland's (1997) theory of occupational choice and previous research (Allen and Robbins, 2008; Nye et al., 2012; Tracey et al., 2012; Le et al., 2014; Le and Robbins, 2016; Nguyen et al., 2016; Kim and Beier, 2020), we found higher interest congruence (i.e., lower Euclidean distance) for students who persisted in their areas and finished their studies successfully. Thus, hypothesis 1 could be confirmed. The effects (differences) of SOC were overall more distinctive than the ones of ASP – especially for male students who didn't show significant effects regarding ASP. These overall differences can be broken down to specific subject areas. Especially for STEM areas, SOC is associated with finishing the studies successfully. In addition, our findings regarding academic performance measures (esp. grades) are in line with the results of recent research (Allen and Robbins, 2010; Tracey et al., 2012; Nye et al., 2018) and meta-analytic findings (Nye et al., 2012) since congruence was substantially related to students' grades. Therefore, hypothesis 2 could also be confirmed. In addition, the results of our study indicate interaction effects as some effects are different for male and female students in different subject areas confirming the conclusions of previous research that the relation between congruence and academic performance is complex and that discipline-unspecific analyzes are insufficient (Nguyen et al., 2016). As with previous research (Assouline and Meir, 1987; Tranberg et al., 1993; Tsabari et al., 2005), the effects of congruence on academic satisfaction indicators were only marginal and weaker than those on persistence and performance. In this regard, having chosen an academic area and aspiring an occupation that fits one's vocational interests has less impact on being satisfied with the study conditions and with the study in general; however, it has more impact on getting better grades.

The stronger impact of congruence on performance than on satisfaction is striking since it not only appears in the academic context but also in the occupational (Hoff et al., 2020).

Distinctive Effects of Congruence

Reflecting on our study's outcomes, we can see distinctive effects of social and aspirational congruence. While SOC was more important for students' persistence, ASP was rather important for students' performance. We would interpret these differences that students who feel a high congruence with their study mates are more inclined to persist within their studies while students that are more identified with their later job aspiration show more performance. The latter effect may be even stronger as ASP could only be evaluated for students that provided an occupational aspiration at study entry. Thus, looking at our study outcomes, the two types of congruence are rather conceptually different aspects of the person-environment fit than equivalent operationalizations of the same construct (c.f. Kristof-Brown et al., 2005; Su et al., 2015). Since SOC indicates the similarity to the other people in a study subject it may considerably impact the current feeling of belonging, while ASP may be more related to one's future goals and may therefore lead to put the focus on the need to perform in order to reach the aspired occupation.

Distinctive Effects for Study Area and Gender

Aside from the differences between the two kinds of congruence, there are area specific and gender differences that are partially responsible for these overall effects. Regarding the gender and subject area, we could see many effects disappearing far beyond significance levels for several areas while the remaining effects increased their effect sizes. This, again, indicates the need for gender and discipline-specific analyses (see e.g., Nguyen et al., 2016).

Person-environment fit theories such as Holland's theory of vocational choice claim that, in general, the fit between a person and an environment is associated with favorable outcomes. The results of the current study indicate that this general statement can be viewed critically insofar as personal factors such as gender can be related to contextual factors such as sociostructural barriers that influence the effect of congruence (Lent, 2013). In the current study, for STEM subjects with its characteristics, congruence seems to be a more important factor than for other subjects. Considering the big efforts for motivating students studying STEM as well as the high dropout rates, it appears that the study characteristics of STEM studies depend much more on congruence than other subject areas. The reason for this could be that for females in STEM – and especially in STEM-L – a good fit to the aspired occupation is especially important in order to be able and willing to overcome barriers such as gender stereotypes or low self-efficacy (Heilman et al., 2004; Le et al., 2014). In this regard, an aspired occupation can serve as a positive outcome expectation influencing performance outcomes (Lent, 2013). However, females still rarely choose STEM areas that are male labeled as vocational aspirations (OECD, 2013). According to Gottfredson (2005), occupations that do not correspond to one's own self-concept due to the wrong sex type are already excluded at the age of 6–8 years, which is why appropriate interventions should be started early. This includes the assumption that the effect of gender on the development of vocational interests and aspirations is mediated by socially constructed processes (Lent et al., 1994).

Furthermore, differences in the SOC can be conveyed to the research on the sense of belonging that is especially a challenge for female students in STEM to obtain (see e.g., Deiglmayr et al., 2019; Höhne and Zander, 2019).

Notably, comparable effects may take place in the languages that show an underrepresentation of male students. In closing, the correlations between congruence and exhaustion at wave 5 may indicate specific assessment situations in medicine and economics at this point in time.

Consequences for Science, Technology, Engineering, Mathematics Interventions

The results have several implications for initiatives for raising young peoples' interest in STEM. First of all, interest congruence is an important factor for study outcomes, persistence as well as performance, in STEM. Therefore, interest assessment and congruence evaluation, e.g., by the Explorix (Jörin Fux et al., 2003), the SDS (Holland et al., 1973), or the O*Net (2018) can be an important starting point for selecting appropriate participants for interventions in two ways: They help to identify participants with a structure of vocational interests congruent to STEM that, however, are not yet motivated for going into STEM. This might be a promising target group for interventions as they also have good chances for persisting in STEM. On the other hand, students with a low congruence in interest assessments may be better counseled toward different pathways as their chances for persisting STEM are notably lower.

The second aspect relates to the different characteristics of congruence. Although ASP is most important for performance, SOC is the key to persistence. Hands-on activities in science labs (e.g., Paechter et al., 2006) or girls' days etc. seem rather to support ASP and students' identification with specific work practices. Such measures seem to motivate students for their later job and rather for performing well in STEM studies. However, such measures tend rather not to cover social aspects like belonging uncertainty (e.g., Deiglmayr et al., 2019; Höhne and Zander, 2019) or stereotypical perspectives about persons working in STEM like Sáinz et al. (2019) report. Thus, initiatives for raising female students' interests in STEM should also focus the social aspect, especially for the STEM-L area that shows a notable under-representation of female students. Here it seems worth thinking about structuring courses toward groups with similar interest profiles for supporting persistence or at least to provide mentoring and/or coaching programs (e.g., Stein, 2013; Kricorian et al., 2020).

The third aspect relates to the perception of being exhausted by the studies that correlates with congruence for STEM areas. These correlations may relate to perceptions of having to be brilliant for being successful in STEM (e.g., Kessels, 2015; Deiglmayr et al., 2019). Stereotypes about the required "brilliance" that is necessary for succeeding in STEM, together with high failure and dropout rates in STEM (see Heublein, 2014) may open a vicious cycle for students with a lower self-concept for STEM. They may invest more effort as they would need up to being exhausted after few semesters in STEM. Rethinking this exhaustion effect from the background of Expectancy-Value-Theory, one can assume that quite a lot of STEM candidates balance their decision against STEM (see Eccles and Wang, 2016), which further supports the stereotype that rather nerds are in STEM areas (see Ertl et al., 2017; Sáinz et al., 2019). This means, on the long run, to rethink STEM curricula, because initiatives to raise participation in STEM are dependent on the perception that a STEM pathway is an attractive career option.

Limitations

As with all other large-scale studies, this study also has several limitations. The first relates to the panel design and the respective panel attrition (see Ertl et al., 2020). Regarding the group of students disappearing, it is unclear how far they completed their studies successfully, failed, or dropped out. Looking into their congruence values in **Table 1**, we could see that these are like the other subgroups. Therefore, we would not expect biasing effects of these students. This applies similarly to students that were so vague in their aspirations that they could neither be classified to a specific occupation nor estimated regarding their ASP. This kind of students may lack in career preparedness (Jaensch et al., 2016) and may be especially present in educational systems of comparably rich countries with low tuition fees like Germany.

Also characteristic for large-scale studies, the current dataset only contained several short scales with comparable low alphas. This issue is intensely discussed in the context of large-scale studies (e.g., Rammstedt and Beierlein, 2014) with Ziegler et al. (2014) concluding that, while less appropriate for individual diagnostics, such short-scales work well for correlational studies like the current one.

Although the strength of this study is to delve deeper into gender- and subject specific differences, this goes along with shrinking sample sizes as well as heterogeneous group sizes for subject areas. This required the use of robust measures that also implied a loss of statistical power. Considering, however, the subject-specific results particularly in **Table 2**, we hope we were able to demonstrate, by providing the p-values, that the non-significances mainly indicate non-effects rather than missing statistical power.

Like this, the study was able to report several significant correlations below the Cohen (1988) threshold of $r < 0.1$. While Cohen (1988) estimates them as marginal or "zero" correlations, and Hattie (2009) postulates only effect sizes larger than 0.4 as desired effects, Funder and Ozer (2019) discuss the value of small and very small effects when they are generalized. This perspective has consequences when considering the almost three Million students and the half Million new students each year in Germany[3]. Although e.g., the correlations between congruence and exhaustion are around $r = 0.05$ which would traditionally mean only 0.25% explained variance, Funder and Ozer (2019) would argue that exhaustion is notably influenced by a missing congruence for several thousand students. We therefore also interpreted several of the low correlations, however, quite cautiously.

CONCLUSION

In our study we found evidence for differences between social and aspirational congruence as well as evidence for differences with respect to gender and subject area. Thus, our study supports the conclusion of Nguyen et al. (2016) that discipline-specific approaches are necessary when analyzing factors for university success and retention. Notably, our results also suggest that gender is an important factor to be considered.

The STEM areas were a specific focus of this study and, in comparison with other study areas, we found that congruence, social as well as aspirational, is especially important in these areas. The study could observe several significant or highly significant yet marginal correlations between congruence and outcome measures. These direct future research toward more detailed analyses including further mediating variables. This aligns well with, e.g., Tracey (2007) who suggested looking deeper into moderators or Kieffer et al. (2004) who proposed applying more complex models. The differences which we found between the subject areas, however, indicate that the respective models might be different and consequently include different variables for the subject areas as well as for gender. Especially for female students in STEM, not only interest congruence but also role congruence may be important factors to consider (see Yang and Barth, 2015). While this study aimed at contextualizing the effects of congruence for a broad range of subject areas, follow-up mediation analyses might only be able to compare models for two or three different subject areas.

Our results also opened conclusions for career counseling and initiatives for raising young people's interest in STEM. Primarily, congruence is an appropriate predictor for study persistence and performance; therefore, it could help to support students' career decisions as well as the selection of participants for STEM programs. However, congruence is for several subjects more important than for other ones, and thus, career counseling should consider that congruence may be less important for some areas (see also Tracey et al., 2012, p. 48). Further research should therefore have a closer look at environmental constraints that contribute to these differences. Our study made a step forward to this approach by distinguishing SOC (fit with the study group) and ASP (fit with the profession aspired) and found differences in the effects of both. The effects of SOC on persistence especially points toward the risk of self-stabilizing systems: When students fit in better with their classmates, they have higher chance of continuing their study; consequently, study areas homogenize. This may especially apply in areas in which students are underrepresented and have a lower sense of belonging like e.g., female students in STEM-L (see Deiglmayr et al., 2019; Höhne and Zander, 2019). Thus, one should consider that interest assessments may propagate gender differences in occupations –

[3]https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/11/PD21_538_21.html;jsessionid=297FC0FD90B067F4E3C4F430121BDCFC.live711

especially in non-traditional ones like Ludwikowski et al. (2019) discuss. Career advisers should be sensitized for such issues, and advice set engines, like Schelfhout et al. (2019a) describe, should be programmed, respectively.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://dx.doi.org/10.5157/NEPS:SC5:14.0.0.

ETHICS STATEMENT

The analyses of this manuscript are secondary analyses of data published previously (Blossfeld et al., 2011). Data sources used for the analyses were the cohort of first year students (doi: 10.5157/NEPS:SC5:14.0.0) of the German National Educational Panel Study (Blossfeld et al., 2011). All students from this cohort gave informed consent to participate in the panel by providing their phone number for being contacted for telephone interviews after being informed about the purposes of the study. Specific information about the recruitment process can be found in the field report of the study (Steinwede and Aust, 2012). All data analyses were performed via a remote terminal (RemoteNEPS) at the LIfBi in Bamberg, Germany that provided a controlled privacy environment for data access. Furthermore, an ethics approval for the analyses was obtained by the local ethics committee.

AUTHOR CONTRIBUTIONS

All the authors contributed to the study conception and design, commented on previous versions of the manuscript, and read and approved the final manuscript. BE and AW performed the material preparation and data analysis. BE and FH wrote the first draft of the manuscript.

FUNDING

This manuscript has been funded by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG), Grant number ER470/2-1.

ACKNOWLEDGMENTS

We would like to thank Diana Lee Sosa for the language revision of this manuscript. This manuscript uses data from the National Educational Panel Study (NEPS; see Blossfeld and Roßbach, 2019). The NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi, Germany) in cooperation with a nationwide net-work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.816620/full#supplementary-material

REFERENCES

Allen, J., and Robbins, S. (2010). Effects of interest–major congruence, motivation, and academic performance on timely degree attainment. *J. Counsel. Psychol.* 57, 23–35. doi: 10.1037/a0019085

Allen, J., and Robbins, S. B. (2008). Prediction of college major persistence based on vocational interests, academic preparation, and first-year academic performance. *Res. High. Educ.* 49, 62–79. doi: 10.1007/s11162-007-9064-5

Assouline, M., and Meir, E. I. (1987). Meta-analysis of the relationship between congruence and well-being measures. *J. Vocat. Behav.* 31, 319–332. doi: 10.1016/0001-8791(87)90046-7

Bai, L., and Liao, H.-Y. (2019). The relation between interest congruence and college major satisfaction: Evidence from the basic interest measures. *J. Career Asses.* 27, 628–644. doi: 10.1177/1069072718793966

Blossfeld, H.-P., and Roßbach, H.-G. (Eds.). (2019). *Education as a Lifelong Process: The German National Educational Panel Study (NEPS). Edition ZfE* (2nd ed.). Springer VS.

Blossfeld, H.-P., Roßbach, H.-G., and von Maurice, J. (2011). Education as a lifelong process: The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft Sonderheft* 14:330.

Brown, S. D., and Gore, P. A. Jr. (1994). An evaluation of interest congruence indices: Distribution characteristics and measurement properties. *J. Vocat. Behav.* 45, 310–327. doi: 10.1006/jvbe.1994.1038

Bruch, M. A., and Krieshok, T. S. (1981). Investigative versus realistic Holland types and adjustment in theoretical engineering majors. *J. Vocat. Behav.* 18, 162–173. doi: 10.1016/0001-8791(81)90004-X

Camp, C. C., and Chartrand, J. M. (1992). A comparison and evaluation of interest congruence indices. *J. Vocat. Behav.* 41, 162–182. doi: 10.1016/0001-8791(92)90018-U

Chartrand, J. M., Camp, C. C., and McFadden, K. L. (1992). Predicting academic adjustment and career indecision: a comparison of self-efficacy, interest congruence, and commitment. *J. Coll. Stud. Dev.* 33, 293–300.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah NJ: Lawrence Erlbaum Associates.

Deiglmayr, A., Stern, E., and Schubert, R. (2019). Beliefs in "Brilliance" and Belonging Uncertainty in Male and Female STEM Students [Original Research]. *Front. Psychol.* 10:1114. doi: 10.3389/fpsyg.2019.01114

Dik, B. J., Strife, S. R., and Hansen, J. I. C. (2010). The flip side of Holland type congruence. Incongruence and job satisfaction. *Career Dev. Q.* 58, 352–358. doi: 10.1002/j.2161-0045.2010.tb00183.x

Donohue, R. (2006). Person-environment congruence in relation to career change and career persistence. *J. Vocat. Behav.* 68, 504–515. doi: 10.1016/j.jvb.2005.11.002

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

Eder, F. (1998). "Differenziertheit der Interessen als Prädiktor der Interessenentwicklung," in *Pädagogisch-psychologische Interessenforschung in Studium und Beruf*, eds J. Abel and C. Tarnai (Münster: Waxmann), 63–77.

Elder, G. H. (2002). "Historical times and lives: A journey through time and space," in *Looking at lives: American longitudinal studies of the twentieth century*, eds E. Phelps, F. F. Furstenberg, and A. Colby (New York, NY: Russell Sage Foundation), 194–218.

Ertl, B., and Hartmann, F. G. (2019). The Interest Profiles and Interest Congruence of Male and Female Students in STEM and Non-STEM Fields. *Front. Psychol.* 10:897. doi: 10.3389/fpsyg.2019.00897

Ertl, B., Hartmann, F. G., and Heine, J.-H. (2020). Analyzing Large-Scale Studies: Benefits and Challenges. *Frontiers in Psychology* 11:3190. doi: 10.3389/fpsyg.2020.577410

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 [Original Research]. *Frontiers in Psychology* 8:703. doi: 10.3389/fpsyg.2017.00703

Etzel, J. M., and Nagy, G. (2016). Students' perceptions of person–environment fit: Do fit perceptions predict academic success beyond personality traits? *Journal of Career Assessment* 24, 270–288. doi: 10.1177/1069072715580325

FDZ-LIfBi (2021a). *Codebook. NEPS Starting Cohort 5—First-Year Students. From Higher Education to the Labor Market. Scientific Use File Version 14.1.0*. Bamberg: Research Data Center LIfBi.

FDZ-LIfBi (2021b). *Data Manual. NEPS Starting Cohort 5—First-Year Students. From Higher Education to the Labor Market. Scientific Use File Version 14.1.0*. Bamberg: Research Data Center LIfBi.

Fu, M., Zhang, L.-F., and Li, B. (2019). Revisiting the congruence–satisfaction relationship: The role of external forces. *J. Career Dev.* 46, 203–218. doi: 10.1177/0894845317737379

Funder, D. C., and Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Adv. Methods Pract. Psychol. Sci.* 2, 156–168. doi: 10.1177/2515245919847202

Gottfredson, L. S. (2005). "Applying Gottfredson's Theory of Circumscription and Compromise in Career Guidance and Counseling," in *Career development and counseling: Putting theory and research to work*, eds S. D. Brown and R. W. Lent (Hoboken, NJ: John Wiley & Sons, Inc.). 71–100.

Hartmann, F. G., Heine, J.-H., and Ertl, B. (2021). Concepts and Coefficients Based on John L. Holland's Theory of Vocational Choice - Examining the R Package holland. *Psych* 3, 728–750.

Hattie, J. (2009). *Visible Learning - A Synthesis of over 800 Meta-analyses relating to Achievement*. Milton Park: Routledge.

Heilman, M., Wallen, A., Fuchs, D., and Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *J. Appl. Psychol.* 2004:416. doi: 10.1037/0021-9010.89.3.416

Henry, P. (1989). Relationship between Academic Achievement and Measured Career Interest: Examination of Holland's Theory. *Psychol. Rep.* 64, 35–40. doi: 10.2466/pr0.1989.64.1.35

Heublein, U. (2014). Student Drop-out from German Higher Education Institutions. *Eur. J. Educ.* 49, 497–513. doi: 10.1111/ejed.12097

Hoeschler, P., and Backes-Gellner, U. (2017). *Shooting for the stars and failing: College dropout and self-esteem* (Economics of Education Working Paper Series). Zürich: Department of Business Administration (IBW), University of Zurich.

Hoff, K. A., Song, Q. C., Wee, C. J. M., Phan, W. M. J., and Rounds, J. (2020). Interest fit and job satisfaction: a systematic review and meta-analysis. *J. Vocat. Behav.* 123:103503. doi: 10.1016/j.jvb.2020.103503

Höhne, E., and Zander, L. (2019). Sources of Male and Female Students' Belonging Uncertainty in the Computer Sciences. *Front. Psychol.* 10:1740. doi: 10.3389/fpsyg.2019.01740

Holland, J. L. (1997). *Making Vocational Choices. A Theory of Vocational Personalities and Work Environments*, 3rd Edn. Odessa: Psychological Assessment Resources.

Holland, J. L., Sorensen, A. B., Clark, J. P., Nafziger, D. H., and Blum, Z. D. (1973). Applying an occupational classification to a representative sample of work histories. *J. Appl. Psychol.* 58, 34–41. doi: 10.1037/h0035405

Hubert, L., and Arabie, P. (1987). Evaluating order hypotheses within proximity matrices. *Psychol. Bull.* 102, 172–178. doi: 10.1037/0033-2909.102.1.172

Iachan, R. (1984). A measure of agreement for use with the Holland classification system. *J. Vocat. Behav.* 24, 133–141. doi: 10.1016/0001-8791(84)90001-0

International Labour Office (2020). *ISCO-08 Structure, index correspondence with ISCO-88*. Nairobi: International Labour Office.

Jaensch, V. K., Hirschi, A., and Spurk, D. (2016). Relationships of vocational interest congruence, differentiation, and elevation to career preparedness among university students. *Zeitschrift Für Arbeits- Und Organisationspsychologie* 60, 79–89. doi: 10.1026/0932-4089/a000210

Jörin Fux, S., Stoll, F., Bergmann, C., and Eder, F. (2003). *EXPLORIX® - das Werkzeug zur Berufswahl und Laufbahnplanung. Deutschsprachige Adaption und Weiterentwicklung des Self-Directed Search (SDS) nach John Holland*. Edison, NJ: Huber.

Juntunen, C. L., Motl, T. C., and Rozzi, M. (2019). "Major career theories: International and developmental perspectives," in *International Handbook of Career Guidance*, eds J. A. Athanasou and H. N. Perera (New York, NY: Springer International Publishing), 45–72. doi: 10.1007/978-3-030-25153-6

Kessels, U. (2015). Bridging the Gap by Enhancing the Fit: how Stereotypes about STEM Clash with Stereotypes about Girls. *Internat. J. Gend. Sci. Technol.* 7, 280–296.

Kieffer, K. M., Schinka, J. A., and Curtiss, G. (2004). Person-Environment Congruence and Personality Domains in the Prediction of Job Performance and Work Quality. *J. Couns. Psychol.* 51, 168–177. doi: 10.1037/0022-0167.51.2.168

Kim, M. H., and Beier, M. E. (2020). The college-to-career transition in STEM: an eleven-year longitudinal study of subjective and objective vocational interest fit. *J. Vocat. Behav.* 2020:103506. doi: 10.1016/j.jvb.2020.103506

Kricorian, K., Seu, M., Lopez, D., Ureta, E., and Equils, O. (2020). Factors influencing participation of underrepresented students in STEM fields: matched mentors and mindsets. *Internat. J. STEM Educ.* 7:16. doi: 10.1186/s40594-020-00219-2

Kristof-Brown, A. L., Zimmerman, R. D., and Johnson, E. C. (2005). Consequences of individuals' fit at work: a meta-analysis of person–job, person–organization, person–group, and person–supervisor fit. *Person. Psychol.* 58, 281–342. doi: 10.1111/j.1744-6570.2005.00672.x

Le, H., and Robbins, S. B. (2016). Building the STEM pipeline: Findings of a 9-year longitudinal research project. *J. Vocat. Behav.* 95–96, 21–30. doi: 10.1016/j.jvb.2016.07.002

Le, H., Robbins, S. B., and Westrick, P. (2014). Predicting student enrollment and persistence in college STEM fields using an expanded P-E fit framework: a large-scale multilevel study. *J. Appl. Psychol.* 99, 915–947. doi: 10.1037/a0035998

Lent, R. W. (2013). "Social cognitive career theory," in *Career development and counseling. Putting theory and research to work*, eds S. D. Brown and R. W. Lent (Hoboken, NJ: John Wiley & Sons), 115–146.

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. Vocat. Behav.* 45, 79–122. doi: 10.1006/jvbe.1994.1027

Leuwerke, W. C., Robbins, S., Sawyer, R., and Hovland, M. (2004). Predicting Engineering Major Status From Mathematics Achievement and Interest Congruence. *J. Career Assess.* 12, 135–149. doi: 10.1177/1069072703257756

Ludwikowski, W. M. A., Schechinger, H. A., and Armstrong, P. I. (2019). Are Interest Assessments Propagating Gender Differences in Occupations? *J. Career Assess.* 28, 14–27. doi: 10.1177/1069072718821600

Mouton, D., Zhang, H., and Ertl, B. (2020). German university student's reasons for dropout: identifying latent classes. *J. Educ. Res.* 12, 190–224.

Muchinsky, P. M., and Monahan, C. J. (1987). What is person-environment congruence? Supplementary versus complementary models of fit. *J. Vocat. Behav.* 31, 268–277. doi: 10.1016/0001-8791(87)90043-1

Nagy, G., Trautwein, U., and Lüdtke, O. (2010). The structure of vocational interests in Germany: Different methodologies, different conclusions. *J. Vocat. Behav.* 76, 153–169. doi: 10.1016/j.jvb.2007.07.002

Neugebauer, M., Heublein, U., and Daniel, A. (2019). Studienabbruch in Deutschland: Ausmaß, Ursachen, Folgen, Präventionsmöglichkeiten. *Zeitschrift für Erziehungswissenschaft* 22, 1025–1046. doi: 10.1007/s11618-019-00904-1

Nguyen, T.-L. K., Williams, A., and Ludwikowski, W. M. A. (2016). Predicting student success and retention at an HBCU via interest-major congruence and academic achievement. *J. Career Asses.* 25, 552–566. doi: 10.1177/1069072716651870

Nichols, R. C., and Holland, J. L. (1963). Prediction of the first year college performance of high aptitude students. *Psychol. Monogr.* 77, 1–29. doi: 10.1037/h0093846

Nye, C. D., Prasad, J., Bradburn, J., and Elizondo, F. (2018). Improving the operationalization of interest congruence using polynomial regression. *J. Vocat. Behav.* 104, 154–169. doi: 10.1016/j.jvb.2017.10.012

Nye, C. D., Su, R., Rounds, J., and Drasgow, F. (2012). Vocational interests and performance. A quantitative summary of over 60 years of research. *Perspect. Psycholog. Sci.* 7, 384–403. doi: 10.1177/1745691612449021

Nye, C. D., Su, R., Rounds, J., and Drasgow, F. (2017). Interest congruence and performance: Revisiting recent meta-analytic findings. *J. Vocat. Behav.* 98, 138–151. doi: 10.1016/j.jvb.2016.11.002

O*Net (2018). *O*NET® 22.3 Database* Version 22.3 [xls].

OECD (2013). *Closing the Gender Gap. Act Now*. Paris: OECD Publishing.

Ott-Holland, C. J., Huang, J. L., Ryan, A. M., Elizondo, F., and Wadlington, P. L. (2013). Culture and vocational interests: the moderating role of collectivism and gender egalitarianism. *J. Coun. Psychol.* 60, 569–581. doi: 10.1037/a0033587

Paechter, M., Jones, G., Tretter, T., Bokinsky, A., Kubasko, D., Negishi, A., et al. (2006). "Hands-on in science education: Multimedia instruction that is appealing to female and male students," in *Multimedia Applications in Education*, eds D. Grabe and H. Zimmermann (Graz: FH Joanneum), 78–85.

Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: missing link between interests and occupations? *J. Vocat. Behav.* 21, 259–287. doi: 10.1016/0001-8791(82)90036-7

Rammstedt, B., and Beierlein, C. (2014). Can't We Make It Any Shorter? *J. Indiv. Diff.* 35, 212–220. doi: 10.1027/1614-0001/a000141

Rounds, J., Smith, T., Hubert, L., Lewis, P., and Rivkin, D. (1999). *Development of occupational interest profiles for O*NET*. Washington, DC: National Center for O*NET Development.

Rounds, J., and Tracey, T. J. (1996). Cross-cultural structural equivalence of RIASEC models and measures. *J. Couns. Psychol.* 43, 310–329. doi: 10.1037/0022-0167.43.3.310

Sáinz, M., Martínez-Cantos, J.-L., Rodó-de-Zárate, M., Romano, M. J., Arroyo, L., and Fàbregues, S. (2019). Young Spanish People's Gendered Representations of People Working in STEM. A Qualitative Study [Original Research]. *Front. Psychol.* 10:996. doi: 10.3389/fpsyg.2019.00996

Sarcletti, A., and Müller, S. (2011). Zum Stand der Studienabbruchforschung. Theoretische Perspektiven, zentrale Ergebnisse und methodische Anforderungen an künftige Studien. *Zeitschrift für Bildungsforschung* 1, 235–248. doi: 10.1007/s35834-011-0020-2

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., De Fruyt, F., and Duyck, W. (2019b). The effects of vocational interest on study results: Student person – environment fit and program interest diversity. *PLoS One* 14:e0214618. doi: 10.1371/journal.pone.0214618

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., De Fruyt, F., and Duyck, W. (2019a). From Interest Assessment to Study Orientation: an Empirical Advice Set Engine. *J. Exp. Educ.* 2019, 1–26. doi: 10.1080/00220973.2019.1605327

Schmitt, N., Oswald, F. L., Friede, A., Imus, A., and Merritt, S. (2008). Perceived fit with an academic environment: attitudinal and behavioral outcomes. *J. Vocat. Behav.* 72, 317–335. doi: 10.1016/j.jvb.2007.10.007

Schneider, M., and Yin, L. (2011). The high cost of low graduation rates: how much does dropping out of college really cost? *Am. Instit. Res.* 2011:1. doi: 10.1037/e537282012-001

Stein, M. (2013). "Von Paten und Lotsen. Coaching- und Mentorenprogramme in der Studien- und Berufsorientierung," in *Berufsorientierung. Ein Lehr- und Arbeitsbuch*, eds T. Brüggemann and S. Rahn (Münster: Waxmann Studium), 271–280.

Steinwede, J., and Aust, F. (2012). Methodenbericht. NEPS Startkohorte 5 - CATI-Haupterhebung Herbst 2010 B52. infas Institut für angewandte Sozialwissenschaften. Available online at: https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/11-0-0/SC5_11-0-0_FieldReports.zip (accessed January 2, 2019).

Su, R., Murdock, C., and Rounds, J. (2015). "Person-environment fit," in *APA handbook of career intervention, Volume 1: Foundations*, eds P. J. Hartung, M. L. Savickas, and W. B. Walsh (Washington, D.C: American Psychological Association), 81–98. doi: 10.1037/14438-005

Tracey, T. J., and Rounds, J. B. (1993). Evaluating Holland's and Gati's vocational-interest models: a structural meta-analysis. *Psychol. Bull.* 113, 229–246. doi: 10.1037/0033-2909.113.2.229

Tracey, T. J. G. (1997). RANDALL: a Microsoft FORTRAN Program for a Randomization Test of Hypothesized Order Relations. *Educ. Psychol. Meas.* 57, 164–168. doi: 10.1177/0013164497057001012

Tracey, T. J. G. (2007). Moderators of the interest congruence-occupational outcome relation. *Internat. J. Educ. Vocat. Guid.* 7, 37–45. doi: 10.1007/s10775-006-9111-3

Tracey, T. J. G., Allen, J., and Robbins, S. B. (2012). Moderation of the relation between person–environment congruence and academic success: environmental constraint, personal flexibility and method. *J. Vocat. Behav.* 80, 38–49. doi: 10.1016/j.jvb.2011.03.005

Tracey, T. J. G., and Sodano, S. M. (2013). "Structure of interests and competence perceptions," in *Handbook of vocational psychology. Theory, research, and practice*, eds W. B. Walsh, M. L. Savickas, and P. J. Hartung (Milton Park: Talyor & Francis), 104–120.

Tranberg, M., Slane, S., and Ekeberg, S. E. (1993). The relation between interest congruence and satisfaction: a metaanalysis. *J. Vocat. Behav.* 42, 253–264. doi: 10.1006/jvbe.1993.1018

Tsabari, O., Tziner, A., and Meir, E. I. (2005). Updated meta-analysis on the relationship between congruence and satisfaction. *J. Career Assess.* 13, 216–232. doi: 10.1177/1069072704273165

Van Iddekinge, C. H., Roth, P. L., Putka, D. J., and Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *J. Appl. Psychol.* 96, 1167–1194. doi: 10.1037/a0024343

von Maurice, J., and Nagy, G. (2009). Kurzskala zur Erfassung der Interessen nach dem Modell von Holland: Das Interesseninventar für die Lebensspanne (IILS) – Kind- und Erwachsenenversion (IILS-I und ILLS-II) (Unveröffentlichter Bericht). Bamberg: Self-published.

Westermann, R., Heise, E., Spies, K., and Trautwein, U. (1996). Identifikation und Erfassung von Komponenten der Studienzufriedenheit [Identifying and assessing components of student satisfaction]. *Psychologie in Erziehung und Unterricht* 43, 1–22.

Wiegand, J. P., Drasgow, F., and Rounds, J. (2021). Misfit matters: A re-examination of interest fit and job satisfaction. *J. Vocat. Behav.* 125:103524. doi: 10.1016/j.jvb.2020.103524

Wohlkinger, F., Ditton, H., von Maurice, J., Haugwitz, M., and Blossfeld, H.-P. (2011). 10 Motivational concepts and personality aspects across the life course. *Zeitschrift für Erziehungswissenschaft* 14:155. doi: 10.1007/s11618-011-0184-5

Yang, Y., and Barth, J. M. (2015). Gender differences in STEM undergraduates' vocational interests: people–thing orientation and goal affordances. *J. Vocat. Behav.* 91, 65–75. doi: 10.1016/j.jvb.2015.09.007

Ziegler, M., Kemper, C. J., and Kruyen, P. (2014). Short Scales – Five Misunderstandings and Ways to Overcome Them. *J. Indiv. Diff.* 35, 185–189. doi: 10.1027/1614-0001/a000148

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: 27 June 2022
doi: 10.3389/fpsyg.2022.937058

On the Design and Validation of Assessing Tools for Measuring the Impact of Programs Promoting STEM Vocations

María Pilar Herce-Palomares [1]*, *Carmen Botella-Mascarell* [2], *Esther de Ves* [2], *Emilia López-Iñesta* [3]*, *Anabel Forte* [4], *Xaro Benavent* [2] and *Silvia Rueda* [2]

[1] International PhD School (EIDUNED), National Distance Education University (UNED), Madrid, Spain, [2] Department of Computer Science, Universitat de València, Burjassot, Spain, [3] Department of Didactics of Mathematics, Universitat de València, València, Spain, [4] Department of Statistics and Operational Research, Universitat de València, Burjassot, Spain

This paper presents the design and validation process of a set of instruments to evaluate the impact of an informal learning initiative to promote Science, Technology, Engineering, and Mathematics (STEM) vocations in students, their families (parents), and teachers. The proposed set of instruments, beyond assessing the satisfaction of the public involved, allow collecting data to evaluate the impact in terms of changes in the consideration of the role of women in STEM areas and STEM vocations. The procedure followed to develop the set of instruments consisted of two phases. In the first phase, a preliminary version (v1) of the questionnaires was designed based on the objectives of the Girls4STEM initiative, an inclusive project promoting STEM vocations between 6 and 18 years old boys and girls. Five specific questionnaires were designed, one for the families (post activity), two for the students (pre and post activity) and two for the teachers (pre and post avitivity). A refined version (v2) of each questionnaire was obtained with evidence of content validity after undergoing an expert judgment process. The second phase was the refinement of the (v2) instruments, to ascertain the evidence of reliability and validity so that a final version (v3) was derived. In the paper, a high-quality set of good practices focused on promoting diversity and gender equality in the STEM sector are presented from a Higher Education Institution perspective, the University of Valencia. The main contribution of this work is the achievement of a set of instruments, rigorously designed for the evaluation of the implementation and effectiveness of a STEM promoting program, with sufficient validity evidence. Moreover, the proposed instruments can be a reference for the evaluation of other projects aimed at diversifying the STEM sector.

Keywords: diversity in STEM, gender stereotypes, informal education, self-efficacy, questionnaire validation, mixed methods

OPEN ACCESS

Edited by:
Sergi Fàbregues,
Open University of Catalonia, Spain

Reviewed by:
Elizabeth G. Creamer,
Virginia Tech, United States
Lina Montuori,
Universitat Politècnica de València, Spain

***Correspondence:**
María Pilar Herce-Palomares
mherce2@alumno.uned.es
Emilia López-Iñesta
emilia.lopez@uv.es

Specialty section:
This article was submitted to Gender, Sex and Sexualities, a section of the journal Frontiers in Psychology

Received: *05 May 2022*
Accepted: *31 May 2022*
Published: *27 June 2022*

Citation:
Herce-Palomares MP, Botella-Mascarell C, de Ves E, López-Iñesta E, Forte A, Benavent X and Rueda S (2022) On the Design and Validation of Assessing Tools for Measuring the Impact of Programs Promoting STEM Vocations. Front. Psychol. 13:937058. doi: 10.3389/fpsyg.2022.937058

1. INTRODUCTION

In recent years, multiple initiatives have emerged, from public and private institutions, to promote interest in disciplines related to Science, Technology, Engineering and Mathematics (STEM), especially among girls from an early age. These initiatives play a fundamental role in showing the relationship that exists between careers and professions in STEM areas and the generation of benefits in society. In addition, they serve to increase the visibility of proximity STEM female referents (UNESCO, 2017), helping to eliminate gender stereotypes (Sáinz et al., 2019).

The School of Engineering of the University of Valencia (ETSE-UV), in Spain, launched in 2011 a pilot program focused on increasing and retaining the number of Information and Communication Technology (ICT) female students in the institution (Botella et al., 2019). The results showed an increase in the proportion of female students in highly male-dominated ICT-related disciplines with a lower proportion of women in general (López-Iñesta et al., 2020). However, it was also observed that a degree such as Chemical Engineering, traditionally with a higher presence of women, showed a constant decrease in female enrollment. This suggested that a continuous effort was needed from educational institutions, public entities, professionals, and families to break the gender diversity gap in STEM (Sáinz and Müller, 2018; López-Iñesta et al., 2020).

The problem of the gender diversity gap in STEM disciplines, and specially in the ICT field, has been considered and analyzed from different perspectives (see Bian et al., 2017; Diekman et al., 2017, 2019; Sáinz and Müller, 2018; Botella et al., 2019; Sáinz et al., 2019; Benavent et al., 2020; López-Iñesta et al., 2020; Ayuso et al., 2021; Gladstone and Cimpian, 2021; Guenaga et al., 2022 and references therein). From these works, aspects such as the influence of gender stereotypes, the effectiveness of using role models, the concept of self-efficacy in STEM or understanding the impact of communal goal processes arise as fundamental factors to be covered by initiatives or programs focusing on pre-university students and aiming at diversifying STEM. There is a second pool of factors related to STEM working environments (i.e., perception of male-dominated environments, lack of work-life balance) which cannot be directly impacted by these type of initiatives. Instead, a large agreement between different social and economical actors should be sought.

In 2019, the Girls4STEM initiative was launched in the ETSE-UV as an evolution of the pilot program. The main feature of the project is that the target audience comprises pre-university students from 6 to 18 years old, as well as their families and teachers (Benavent et al., 2020). It is a project for both boys and girls, with an emphasis on girls, which is framed in the Sustainable Development Goals (SDGs) of United Nations and it is also aligned with the III Equality Plan of the University of Valencia (López-Iñesta et al., 2020). The specific objectives of the Girls4STEM initiative are: i) To awake curiosity about STEM disciplines from an early age; ii) To encourage the participation of students, teachers, families, and companies as a fundamental part of the project; iii) To give visibility to women developing their professional work in STEM areas and show their research, developments and progress; and iv) To increase the number of students in STEM studies through outreach activities such as seminars, workshops or interviews with leading women in STEM. The initiative is arranged around two main activities, Girls4STEM *family*, focused on pre-university students, their families and teachers, and Girls4STEM *Professional*, targeting a general audience. Note that a full description of the initiative can be found in Benavent et al. (2020). The initiative builds upon a large database of volunteer female STEM professionals, which are the ones interacting with the students and teachers via the *family* action or with the general audience via the *professional* action. The female STEM professionals act then as proximity role models, mitigating the impact of gender stereotypes, while the database helps increasing the visibility of their contributions to the society, reinforcing the link with communal goal objectives (Botella-Mascarell et al., 2021). In the *family* action, students gather with the STEM experts and they create 3 min videos about them which are later uploaded into the Girls4STEM YouTube channel. A contest is then arranged between the participating schools, where the Girls4STEM initiative selects the videos which best reflect the aims of the project.

The Girls4STEM initiative has been consolidated in two editions, being the edition 2021–2022 currently on-going. At this point, it is essential to have instruments with sufficient evidence of validity to evaluate with scientific rigor the impact of the initiative, as indicated by Tena Gallego and Couso (2019), beyond the satisfaction of the public involved. With this aim, this paper presents the design and validation process followed to obtain a set of instruments to evaluate the impact of the Girls4STEM initiative in the *family* action. To this end, the role of formal and informal learning contexts in STEM education is reviewed next, and the focus is then placed in informal education initiatives.

1.1. State of the Art

STEM education takes place in both formal and informal contexts and both need to be connected to promote students' STEM skills. Interestingly, informal education can overcome many of the shortcomings of formal education (Herce Palomares et al., 2022). Activities promoted by different initiatives or entities such as universities, museums, science fairs or contests are examples of informal education scenarios in which students, teachers, families or citizen participation is promoted (López-Iñesta et al., 2022). The audience and researchers/professionals in different fields can establish a useful bidirectional communication for fostering interest in STEM areas. From this point of view, the Girls4STEM initiative can be classified as an informal education/learning action organized by a Higher Education Institution. Girls4STEM builds bridges with formal education, involving both teachers and students' families from a systemic, integral and holistic educational vision. Although the word "informal" suggests insufficient correctness, it is actually highlighting the features of the learning environment. As pointed out in Allen and Peterman (2019), informal learning might contribute to achieve high levels of area-specific expertise for motivated student's. In addition, research suggests that educational experiences to promote STEM expertise in informal education play a decisive role (Herce Palomares and Román González, 2021) and, they also contribute to challenge common ideas and beliefs linked to STEM fields in formal education, as well as others related to scientific education (Benavent et al., 2020). In informal education learning, evaluation is one of the key components. Whilst helping to identify if aims and objectives have been met, it can also assist with planning, provide evidence of impact, and critically reflect for future engagement activities. Therefore, evaluation is a process that should run from the start of a project and continue after it has finished (Robinson and Murray, 2019).

Evaluating the impact in informal learning contexts poses a set of particular challenges (Habig, 2020). Firstly, evaluations should preserve the informal nature of science experiences, while defining appropriate evaluation metrics, using a common language, goals, and theories (National Research Council, 2009). Coupling these challenges with constraints on time, money, and operational capacity, the difficulty of obtaining meaningful, reliable and feasible evaluations becomes clear. The evaluation should then tackle these challenges to provide useful evidence-based information (Fu et al., 2016). Secondly, formal learning experiences are primarily intended to impart scientific knowledge and skills. However, informal learning experiences are intended to arouse curiosity, interest and encourage intrinsic motivation as "stepping stones" for STEM learning. This increases the difficulty of the evaluation process, since constructs such as interest, motivation and curiosity are more difficult to define, operationalize and measure (National Research Council, 2009). In this sense, evaluating the impact of educational interventions in informal STEM education requires the design of instruments that address the project objectives.

Three future directions for the measurement of the outcomes of informal STEM education actions are suggested in Grack Nelson et al. (2019). First, the measurement capacity should be enhanced. Currently, there is a small number of online repositories, covering also a limited range of activities and audience. Second, stronger collaborative networks should be established. These type of networks would allow to achieve shared measures combining different expertise (measurement experts, educational researchers, STEM experts). Finally, it is mandatory to increase the accessibility of shared measures. There are barriers related to intellectual property rights or instruments not accessible due to journal publishing options.

Another challenge related to the evaluation of the impact in informal STEM education is the broad range of projects and the large variety of methods used to conduct the evaluation. The most common form of evaluation is the user survey (Robinson and Murray, 2019). When designed well and interpreted appropriately, self-report surveys can be used to gather useful data from large samples at relatively low-cost (Wolf et al., 2021). Note that informal education initiatives are usually constrained by low budgets and hence, sustainable implementations should be sought. Therefore, in this work, the user survey technique via questionnaires is proposed to evaluate the impact of the Girls4STEM initiative in the *family* action, by designing and validating a set of questionnaires targeting pre-university students, their families and teachers.

With the increasing development and use of shared measures across the STEM education field, it comes the need for evaluators to better understand and assess instrument's technical qualities, in particular reliability and validity (Grack Nelson et al., 2019). On the one hand, the design of the evaluation instruments must be based on the objectives of the project. However, the questionnaires must undergo a validation process. Content validity evidence relates to how well the construct of interest is represented in the content of an instrument (Haynes et al., 1995; AERA, 2014). Such evidence can be collected by reviewing the literature and gathering feedback from experts related to the construct being measured. Experts review how the construct was defined, identify what is missing from the definition, and help to ensure that the essence of the items or tasks in the measure adequately cover the content area. On the other hand, evidence of the reliability of the questionnaires, after being administered to a pilot sample, is needed. Cronbach's alpha is commonly used to examine the internal consistency or reliability of summated rating scales (Cronbach, 1951; Cronbach and Shavelson, 2004; AERA, 2014), although there is an on-going discussion regarding its limitations (Trizano-Hermosilla and Alvarado, 2016; Xiao and Hau, 2022). Internal consistency describes the extent to which all the items in a test measure the same concept (or construct) and hence it is connected to the inter-relatedness of the items within the test. In addition to obtaining the reliability of the scale items, it is necessary to evaluate how open-response items work in the pilot sample. In this way, it is possible to check whether the answers given in the questionnaires have the same meaning for the target audiences as for the researchers interpreting the data (Wolf et al., 2021). **Figure 1** summarizes the main advantages and challenges faced by STEM informal learning contexts, as well as the main constructs to measure and some hints about the instruments design.

1.2. The Present Study

This study tackles good practices focused on promoting gender diversity in the STEM sector from a Higher Education Institution perspective. A high-quality example of a gender-based intervention study in informal STEM education is presented, with sufficient evidence of the validity of a set of rigorously designed instruments for the evaluation of the implementation and effectiveness of the project. In addition, these instruments can be a reference for the evaluation of other projects aimed at reducing the gender diversity gap in STEM areas. The process and the results presented in this paper contribute to the directions suggested by (Grack Nelson et al., 2019), since the measurement capacity is increased, the questionnaires are accessible to other researchers and hence, there is potential to build a collaborative network. The main objective of this work is then to design and obtain evidence of reliability and validity of a set of instruments designed to evaluate the impact of the Girls4STEM initiative. This objective can be broken down into a set of specific objectives:

1. To design a set of questionnaires to evaluate the impact of the Girls4STEM initiative (*family* action). Each questionnaire will be specific for a different audience group: pre-university students, their families and teachers.
2. To obtain evidence of content validity of the set of questionnaires.
3. To obtain evidence of reliability of the set of questionnaires after administration to a sample and to assess whether the answers in self-assessment questionnaires have the same meaning for the target audiences and the researchers who interpret the data.

As discussed in the introduction, the gender diversity gap in STEM has been already considered from different perspectives. In Spain, the percentage of enrolled female students in the different STEM disciplines is not uniform. For example, in

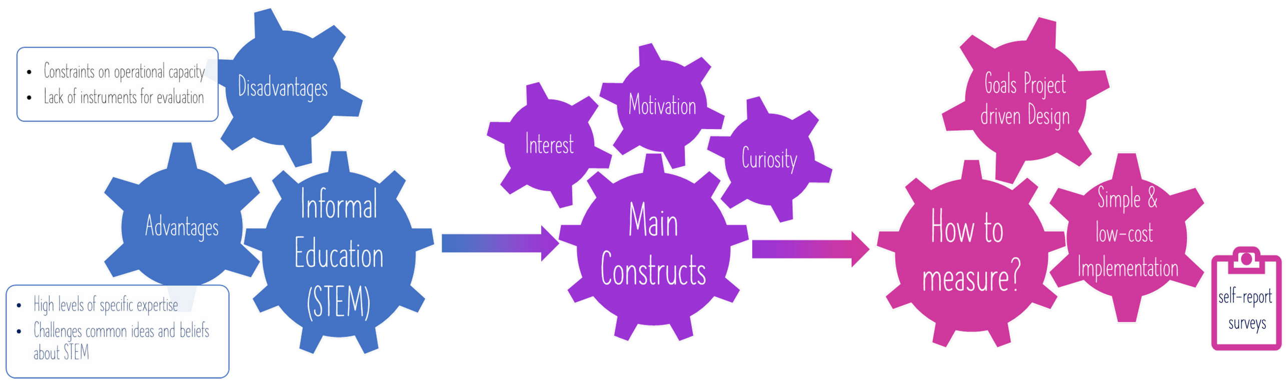

FIGURE 1 | Advantages and disadvantages faced in STEM informal education, main constructs to measure and some hints about the evaluation.

2020-2021, there is a percentage of enrolled female students of 59.9% in life-sciences. In the case of Engineering, the number of enrolled female students goes down to 26.1%, and to 14.2% in the case of Computer Science[1]. There are several initiatives or projects located in Spain that work toward diversifying the STEM sector (Botella et al., 2020). Most of them can be classified as informal education actions, and they also face the evaluation challenges discussed above. Note that some of these initiatives are nodes from international projects. Some representative examples in Spain are, first of all, the Inspira STEAM Program, which is a mentoring program for students between the ages of 10 and 12 years. Results of the program showed an impact on the students' attitudes toward technology, an increase in the number of female STEM referents the student's knew, and an improvement of the students' opinion regarding vocations and professions related to science and technology. Moreover, a larger impact was measured among girls (Guenaga et al., 2022). Secondly, the program by the Inspiring Girls Foundation focuses on pre-university 12–16 years old girls, which interact with female role models working in STEM fields. Reference (González-Pérez et al., 2020) shows a set of benefits on mathematics enjoyment, importance attached to math, expectations of success in math, and girls' aspirations in STEM, and a negative effect on gender stereotypes, among others. Thirdly, the project Science and Technology as Feminine aims at students in the 1st to 3rd years of compulsory secondary education (therefore aged 11–14 years). Results in Santos et al. (2021) show that it should be possible to reduce the gender gap in the future career choices of young students, through the design of a set of activities addressed to individual students, the students' families and peers, schools and society at large, aimed at changing the habits, which for many years have steered women away from STEM. Despite the relevance and impact of the above STEM education initiatives, there is a lack of instruments with evidence of reliability and validity to assess the impact of the projects themselves, since they either make use of questionnaires to measure specific dimensions (i.e., gender stereotypes (Colás Bravo and Villaciervos Moreno, 2007), mathematical self-efficacy (Schwarzer and Baessler, 1996) and attitudes toward technology (Kier et al., 2014)) or questionnaires without a sufficient design and validation process. To the best of our knowledge, this paper contributes to the state of the art of informal STEM education by providing the description of the process and evidences of reliability and validity of a set of instruments that were designed to specifically assess Girls4STEM's objectives.

The paper is organized as follows. Section 2 presents the two phases followed for the design and validation of the proposed set of instruments. Details about the samples used in each one of the phases are given and the data analysis approach followed is explained. The section finishes providing the results obtained in terms of content validity and reliability for the set of instruments. Finally, section 3 discusses the main findings of this research.

2. MATERIALS AND METHODS

The present work uses a Mixed Methods Research (MMR) approach whereby both qualitative and quantitative data are collected and analyzed in the same study. MMR is often used in social and behavioral studies, such as education or health, to strengthen the reliability of qualitative data, allowing to put quantitative results in a context and enriching the findings and conclusions (Creswell and Clark, 2003; Onwuegbuzie and Johnson, 2006; Anguera et al., 2012; Fàbregues et al., 2019). In the specific context of this work, using mixed methods can both increase the validity and reliability of the data collected with the designed instruments and improve the evaluation procedure to measure the impact of the initiative (Shekhar et al., 2019; Griffiths et al., 2021; Hargraves et al., 2021. In this sense, the aim of the study is to design and validate a set of different instruments for measuring the impact on students, parents and teachers of a program promoting STEM vocations that can be used on a large scale by other researchers.

The procedure consisted of two phases. First, in phase I, a preliminary version of the questionnaires was designed by the leading researcher based on the objectives of the Girls4STEM initiative, obtaining a first version (v1) of each one. Afterwards,

[1]Ministerio de Universidades. Students statistics. https://bit.ly/3yA6Bcs.

6 experts participating in the project and with experience in instrument construction and validation, modified and/or polished the items of the different questionnaires through an expert judgment process to obtain evidence of content validity, deriving the version (v2). In the second phase, phase II, the version (v2) instruments were distributed to a pilot-sample. Evidence of reliability was gathered and a final refinement process was carried out. Finally, the final version (v3) was obtained. All the questionnaires collected socio-demographic information and some indicators with a response format with open-ended, multiple choice answers and Likert scale options (1 to 5). **Figure 2** summarizes the steps followed during the process of design and validation of the instruments.

2.1. Instrument: Design and Validation Process

In this subsection, the two-phase process for obtaining the instruments is detailed. Note that there are a total of five questionnaires targeting different groups: parents (post-activity), students-pre (prior to activity), students-post (post-activity), teachers-pre (prior to activity) and teachers-post (post-activity). The first instrument is a questionnaire for families, administered once the participation in the project is finished. It includes indicators on the overall impact of the initiative and on the individual (family member). An indicator is also provided on the possible improvement of the project and the promotion of STEM within the family. Secondly, there are two questionnaires for students that are applied before and after participating in the project. The pre questionnaire collects indicators on STEM interests, their perception of STEM competence and performance in STEM subjects. The post collects indicators on the degree of participation, the impact and possible improvement of the project. The teachers' questionnaires are also arranged in pre and post. The pre includes indicators on motivation and expectations of the project. The post questionnaire asks about their participation degree, the project impact, and suggestions for improvement.

Phase I. Design and evidence of content validity using the expert judgment method. The first phase consisted of two parts. Firstly, an initial version (v1) of the questionnaires was designed by the leading researcher and secondly, evidence of content validity using the expert judgment method was obtained, after which a new version (v2) of each of the five questionnaires was available.

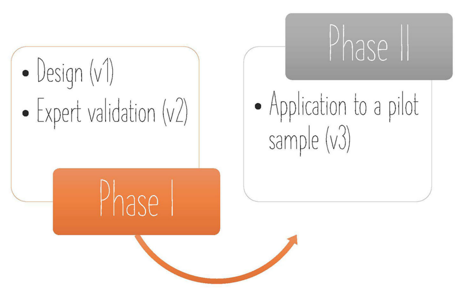

FIGURE 2 | Phase I and phase II stages, and questionnaire versions obtained in each one of them.

The five questionnaires in their initial version (v1) were designed using as a reference the objectives of the Girls4STEM initiative. A set of items was generated to collect inputs from the subjects participating in the *family* action (families/parents, students and teachers), and the dimensions to be measured according to the objectives were specified. An *ad hoc* questionnaire for each of the five questionnaires was then prepared, which was distributed to the committee of experts for undergoing the expert judgment process. These *ad hoc* questionnaires asked about the pertinence/representativeness (whether the items are representative of the dimensions they are intended to measure), relevance (whether the items contribute with important information to the measurement of the dimension) and formulation (whether the items are understood, unambiguous and clear), all on a Likert scale from 1 (not at all in agreement) to 6 (totally in agreement).

In addition, after each set of items, suggestions were requested in open-ended questions when not in complete agreement and an open-ended question was provided at the end of each questionnaire, for any relevant considerations on the design of the instrument. The five *ad hoc* questionnaires were distributed to the committee of experts online, and they were sent to them as well in advance, so that the five questionnaires could be accessed before making their judgments.

Phase II. Distribution of the instruments to a pilot sample. In the second phase, the five instruments in version (v2) were administered through non-probabilistic purposive sampling to a pilot sample of families (parents), students and teachers participating in Girls4STEM in the 2020–2021 academic year. Before the start of the project and the distribution of each questionnaire, informed consent was requested and the current legislation on data protection was complied with, while maintaining the confidentiality of the data. A double analysis (quantitative and qualitative) was performed with the results. First, with the quantitative information, the reliability as internal consistency was calculated from the two-factor model based on the average correlation between the items, using the SPSS v27 program (George and Mallery, 2010), and studying the items on a Likert scale. Secondly, the open-ended questions were analyzed by the group of researchers by means of a content analysis to determine how the questionnaire worked in the population and to be refined if necessary.

2.2. Sample

In this subsection, a description of the sample of each one of the phases is provided.

Phase I. Six female researchers made up the committee of experts. This is a non-probabilistic purposive sample, all of them being women. The selection meets the criteria proposed by Skjong and Wentworth (2000) for purposive sampling: experience in making judgments and decisions based on evidence or expertise, reputation in the community, availability and motivation to participate, impartiality and inherent qualities such as trustworthiness and adaptability.

Phase II. A total of 8 schools, all of them located in the Valencian Community, participated in the Girls4STEM initiative during the 2020–2021 academic year. From these schools, 6 were public and 2 were charter schools. Regarding their geographical origin, 2 of them were located in small cities (*population* $< 30,000$), 3 in medium-sized cities (*population* $< 100,000$), while 3 were located in large cities (*population* $> 100,000$). This brings the total group of students participating to 298, distributed between 84 in small cities, 109 in medium-sized cities and 105 in large cities.

The final sample used for this study, eliminating those students who did not fill in the pre or post questionnaires, was 268 students, 18 teachers (16 female and 2 male teachers) and 113 family members (88 female and 25 male). Therefore, the sample was constructed by non-probability purposive sampling.

Table 1 shows the distribution of participating students according to gender, with a higher percentage of female students (62%), and educational level, defining the following levels: primary, secondary with 2 subgroups by age, and professional studies.

Regarding the education level, the table shows that the largest group was secondary education with students between 12 and 16 years old, accounting for 78% of the total sample. The educational level with the lowest representation in our sample corresponded to secondary education, aged 17–18 (0.03%).

TABLE 1 | Number of students who completed the pre and post questionnaires by gender and educational level.

	Gender			
Educational level	**Male**	**Female**	**Undeclared**	**Total**
Primary	16	15	1	32
Secondary (12–16 years old)	74	135	1	210
Secondary (17–18 years old)	5	4		9
Professional studies	5	12		17
Total	100	166	2	268

2.3. Data Analysis

Data have been processed according to the specific objectives of the research and the established phases. A description of the process followed in each phase is included in this subsection.

Phase I. The SPSS version 27 software was used to calculate the evidence of content validity. Firstly, the mean of the items of each questionnaire in the three dimensions under evaluation (representativeness, relevance, and formulation) was obtained. Given that the Likert scale consisted of 6 points, the criterion for refining an item was that a mean less than 5 were obtained (a value of 5 suggested agreement and 6 suggested total agreement). Secondly, the internal consistency of the judgments issued was calculated by obtaining Cronbach's alpha as intraclass correlation coefficients, according to the bidirectional random model of consistency suggested by Gwet (2014). Finally, the Content Validity Ratio (CVR) of each item was calculated by applying the model of Lawshe (1975) modified by Tristán-López (2008):

$$\mathrm{CVR} = ne/N,$$

where *ne* is the number of experts who gave a favorable judgment (5 or 6 in representativeness) and *N* is the total number of experts who responded to the *ad hoc* questionnaire. The CVR provides evidence of content validity for each indicator. From this model, items are considered essential when scores of 5 and 6 are obtained on the Likert representativeness scale. Any item with a score lower than 0.58 should be deleted (Tristán-López, 2008). The *ad hoc* questionnaires also offered open-ended questions to complete the assessments. In the event that an item needed to be refined, it was modified according to the suggestions of the experts.

Phase II. The data collected after the administration of the version (v2) instruments to a pilot sample of subjects (parents, teachers and students) was analyzed. With the quantitative information (Likert scale questions), Cronbach's alpha reliability coefficient was calculated. With the qualitative information, a content analysis was conducted, in order to assess the performance of the instruments in the sample and to refine them if necessary. Groenvold et al. (1997) suggests that, although rarely investigated, it is necessary to check whether the answers in self-assessment questionnaires have the same meaning for the target audiences as for the researchers who interpret and report the data.

3. RESULTS

This section presents the results of the design and debugging process of the five questionnaires. Results of phase I provide evidence of content validity after the design process, for each of the five questionnaires. Results of the phase II include evidence of reliability of the scale items and an analysis of the performance of the qualitative items.

3.1. Phase I

First, the results related to the specific objectives 1 and 2 of the paper are presented. **Table 2** summarizes the questionnaires in version (v1) including the dimensions, items and scale used in each one. The questionnaires collected the information that was considered appropriate for the measurement of the initiative's objectives, although for the objective of *increasing the number of students in STEM studies*, an indirect measurement of the results was proposed, by assessing interest at the time of the evaluation. As it can be seen in the table, the evaluation was not limited to measuring participant satisfaction. For each set of participants, the measurement of those aspects that were considered critical was proposed. In addition, indicators were included on issues relevant to achieving the aims of Girls4STEM, which are intended to be analyzed in further research, such as family involvement

TABLE 2 | Design of the questionnaires (v1).

Questionnaire	Dimensions (item number)	Scale
Parents	Overall impact (1–3)	2 multiple choice
		1 dichotomous (with open-ended question)
	Impact on parents (4–7)	4 Likert (1–5 points)
	Satisfaction and project improvement (8–10)	1 Likert (1–5 points)
		2 open-ended questions
Students-pre	STEM interests (1–2)	1 dichotomous (with open-ended question)
		1 Likert (1–5 points)
	Achievement in STEM subjects (3)	1 open-ended question
Students-post	Degree of participation (1–2)	2 open-ended questions
	Impact on students (3–6)	4 Likert (1–5 points)
	Satisfaction and project improvement (7–9)	1 Likert (1–5 points)
		2 open-ended questions
Teachers-pre	Motivation toward the project (1–2)	2 open-ended questions
	Expectations (students) (3–5)	3 open-ended questions
	Expectations (teachers) (6)	1 open-ended questions
Teachers-post	Degree of participation (1–2)	2 open-ended questions
	Impact on students (3–5)	3 open-ended questions
	Impact on teachers (6–13)	1 open-ended question
		1 multiple choice
		6 Likert (1–5 points)
	Satisfaction and project improvement (14–15)	1 Likert (1–5 points)
		2 open-ended questions

Dimensions and items, and scale of each one are included in the second and third row, respectively.

in promoting STEM interests, factors that contribute to student involvement in STEM studies, such as achievement or interest (UNESCO, 2017), or the role of teachers in promoting STEM vocations. Note that the questionnaires collected information on socio-demographic data, which is out of the scope of this study.

After the design of the questionnaires in their initial version (v1), the questionnaires were subjected to expert judgment to reach evidence of content validity and to refine the questionnaires, if necessary. *Ad hoc* questionnaires were distributed for expert judgment, and the obtained results for the inter-rater reliability (Cronbach's alpha) are summarized in **Table 3**. In the following, the evidence of content validity is discussed for each questionnaire, both considering the mean of the items of each questionnaire and the internal consistency of the judgments.

TABLE 3 | Inter-rater reliability (Cronbach's alpha).

Questionnaire	Dimension	Cronbach's alpha
Parents	Representativeness	0.262
	Relevance	0.406
	Formulation	0.895
Students-pre	Representativeness	0.8
	Relevance	0.6
	Formulation	0.944
Students-post	Representativeness	0.987
	Relevance	0.981
	Formulation	0.273
Teachers-pre	Representativeness	0.935
	Relevance	0.946
	Formulation	0.359
Teachers-post	Representativeness	0.69
	Relevance	0.92
	Formulation	0.942

Evidence of content validity of the parents questionnaire. The questionnaire for parents (v1) consisted of a total of 10 items (see **Table 2**). The results in terms of the mean of the items after the expert judgment are shown in **Table 4**. In the dimension of representativeness, the mean of all the items ranged between 5.67 and 6, so none of them had to be modified, according to the criterion defined beforehand. Cronnbach's alpha coefficient in **Table 3** suggested sufficient consistency with a value of 0.262. Finally, the CVRs for all the items were 1, which leaded to the conclusion that the questionnaire had sufficient evidence of content validity in the representativeness dimension, i.e., the items were representative of the dimensions they were intended to measure. In the relevance dimension, the results were similar to the ones in the representativeness dimension, with means between 5.67 and 6 (**Table 4**) and a Cronbach's alpha as intraclass correlation of 0.4 (**Table 3**). The formulation

TABLE 4 | Mean (parents).

Mean	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Representative	5.83	6	5.83	6	5.67	6	6	5.67	6	6
Relevance	5.83	6	5.67	6	5.67	6	6	6	6	6
Formulation	6	4	5.83	5.83	4.17	6	6	6	6	6

dimension pointed in another direction. Both item 2 and 4 showed values below 5, so both needed to be reformulated. In spite of this, this dimension presented a high consistency, since Cronbach's alpha value was 0.895. In order to proceed with the refinement, the open-ended questions were analyzed qualitatively. In item 2, two experts suggested introducing "in his/her family" and in item 5, replacing "the role" with "participation." The suggestions were accepted and both items were reformulated.

Evidence of content validity of the students-pre questionnaire. The initial student questionnaire (v1) consisted of three items (see **Table 2**), although the first item offered a dichotomous response which, if affirmative, required an explanation in an open-ended question. **Table 5** shows the results of the mean of the items after the expert judgment for each dimension. In the representativeness dimension, the mean of the items ranged between the values 5.33 and 5.67 (no rephrasing of any of the items necessary). These results were consistent with a Cronbach's alpha of 0.8 (**Table 3**). In addition, none of the items needed to be deleted in terms of the CVR criterion, since all of them reached the maximum value (CVR = 1, except item 2 with CVR = 0.83, which also exceeded 0.58). In the dimension of relevance, Cronbach's alpha (**Table 3**) again suggests consistency in the judgments (Cronbach's alpha = 0.6). The means were higher than in the previous dimension, with values between 5.67 and 6. However, as in the questionnaire for families, the dimension of formulation showed a very high consistency (Cronbach's alpha = 0.944 in this case), but the means indicated the need to reformulate item 2 (mean = 4.17) and 3 (mean = 4.5). Therefore, the open-ended questions of the expert judgment that explained this result were studied. Given that in the Spanish educational system the subjects in the primary and secondary education stages related to STEM contents are different, the experts proposed to specify the term "STEM" in the curricular subjects of both indicators and to not limit the answers to primary education subjects. For version (v2) of this questionnaire, STEM interests (item 2) and school performance (item 3) were defined on the basis of these subjects. Finally, in the open-ended questions at the end of the *ad hoc* questionnaire, it was suggested to incorporate a new dimension, the self-efficacy (perceived achievement), as the experts judges considered it to be a relevant indicator in STEM education. A new indicator was added as requested by the experts.

Evidence of content validity of the students-post questionnaire. The final student questionnaire (v1) consisted of 9 items (see **Table 2**). **Table 6** shows the results of the mean of the items after the expert judgment en each dimension. In the dimension of representativeness, item 1 and 2 were below the criterion (at least 5). In addition, the results showed a high consistency (Cronbach's alpha = 0.987) and the CVR warned about a low-content validity of the first two items, since CVR = 0.33 and CVR = 0.17 for item 1 and 2, respectively. This indicated that both items should be deleted. The experts' feedback on the open-ended questions was reviewed. In item 1, they considered that it was not a decision for the students to take, so the item was not appropriate. For item 2, both in this dimension and in relevance, they suggested incorporating the measurement of the degree of participation with new indicators such as justifying participation in the specific project, and quantitatively specifying the degree of participation in number of hours. Items 1 and 2 were eliminated and two new items were created to evaluate the degree of participation. The information obtained in the results for relevance was similar to the representativeness dimension, with the first two items of the degree of participation being the ones that need to be modified. The means of the items 1 and 2 were again below the criterion. Cronbach's alpha reached a high value (Cronbach's alpha = 0.981) and the open-ended questions raised the point found in the dimension of representativeness. Both items 1 and 2 were reformulated. The formulation dimension showed much more satisfactory results, as all the means were above the criterion and Cronbach's alpha = 0.273, so the consistency was sufficient. No item was subject to change after the results in the formulation. However, in the open-ended question of the final part of the *ad hoc* questionnaire, two experts suggested changing the order of presentation of items 4 and 5. They argued that item 5 was related to the interests raised in item 3, although in this case in relation to the professions. The suggested change in the presentation format was included.

Evidence of content validity of the teachers-pre questionnaire. The initial teacher questionnaire (v1) consisted of 6 open-ended questions items (see **Table 2**). **Table 7** shows the results of the mean of the items after the expert judgment for each dimension. In the representativeness dimension, only item 2 was below the criterion and needed refinement. The inter-rater reliability was sufficient (Cronbach's alpha = 0.935), but item 2 showed a CVR = 0.33, which indicated that the item should be removed from the questionnaire. Item 4 showed a CVR = 0.66, but it was kept in the questionnaire since it exceed the criterion of 0.58 (Tristán-López, 2008). In the dimension of relevance, item 2 was also below the criterion. The judgments were consistent, since Cronbach's alpha = 0.946. Finally, the formulation dimension

TABLE 5 | Mean (students-pre).

Mean	Item 1	Item 2	Item 3
Representative	5.67	5.33	5.33
Relevance	5.67	5.67	6
Formulation	6	4.17	4.50

TABLE 6 | Mean (students-post).

Mean	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9
Representative	3.17	3	6	5.83	5.67	6	6	6	6
Relevance	2.50	2.83	5.83	6	5.50	6	5.67	5.83	5.83
Formulation	5.83	6	5.67	5.83	6	6	6	6	6

TABLE 7 | Mean (teachers-pre).

Mean	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Representative	5.67	3.17	5.67	5	5.67	6
Relevance	5.67	3	5.67	5	5.5	6
Formulation	5.83	5.50	6	6	6	5.83

did not require modification, since the means were above the criterion and Cronbach's alpha = 0.359. In summary, item 2 was eliminated, and version (v2) was composed of 5 items.

Evidence of content validity of the teachers-post questionnaire. The final teacher questionnaire (v1) consisted of 16 items (see **Table 2**). **Table 8** shows the results of the mean of the items after the expert judgment for each dimension. In the dimension of representativeness, item 2 was below the criterion. The results were consistent with Cronbach's alpha = 0.69. The CVR of all the items was 1, except for item 2 where CVR = 0.66. Since this value exceeded the criterion of 0.58, the item did not need to be removed. The results in the dimension of relevance were larger, but item 2 was below the criterion. The judgments were consistent with a Cronbach's alpha = 0.92. In the formulation dimension, the results were similar to the other dimensions, with a Cronbach's alpha = 0.942 and the mean of item 2 below the criterion. The judges open-ended responses were revised for item 2. The suggestion was to divide item 2 and quantify it. Hence, the item 2 ("How much time have you spent on it and how much time have your students spent on it?") was divided in two new items: "Indicate the number of hours you have spent" and "Number of videos in which you have participated." After the modification, version (v2) was composed of 17 items.

Once phase I was completed, all five questionnaires were available in version (v2), with sufficient evidence of content validity in all of them.

3.2. Phase II

In order to collect data for phase II of this study, the pre-questionnaires were administered to students and teachers before interacting with the STEM experts, so gender and professional career aspects have not yet been discussed. The post-questionnaires for students, teachers and families (parents) were administered after each school submitted the STEM expert biography video to the initiative. All the questionnaires were delivered using the Microsoft forms platform. In the following, results related to the specific objective 3 of the paper are analyzed both quantitatively and qualitatively.

3.2.1. Evidence of Reliability

The aim was to ascertain the evidence of reliability and to refine the questionnaires if necessary. To this end, the results were analyzed quantitatively. **Table 9** summarizes the dimensions, scale and analysis type of the different version (v2) questionnaires. The quantitative information was used to determine the evidence of reliability. To this end, reliability was calculated as internal consistency (using the SPSS v27 program), from the two-factor model based on the average correlation between the items that were formulated using a Likert scale. As it can be seen in **Table 9**, this analysis was feasible for all the questionnaires except for the teachers-pre case. **Table 10** shows the results of evidence of reliability for each one of the questionnaires. The second column indicates the number of items that were evaluated (formulated using a Likert scale), the third column stands for the number of valid samples used out of the total number of responses collected from the pilot sample and the fourth column gives the value of the Cronbach's alpha coefficient. In the fifth column, the evaluated item number is provided, while column 6 shows the total correlation of the corrected item and finally, column 7 gives the Cronbach's alpha coefficient if the item is deleted. Note that item 11 of teachers-post questionnaire did not offer results after its calculation, since the answers of all the subjects presented the same value, in this case 5.

George and Mallery (2010) suggest that, in order to evaluate the values of Cronbach's alpha coefficients, a value above 0.7 is considered acceptable. Loewenthal and Lewis (2001) warns that, in scales with less than 10 items, an internal consistency value of 0.6 can be considered acceptable. Results in **Table 10** show that sufficient evidence of validity was achieved in all the questionnaires in the sample used, except for the students-pre questionnaire, with a Cronbach's Alpha = 0.49, which is a low value. The study of the corrected item-total correlation pointed out that item 2B presented a low linear correlation between this item and the total score of the scale. Moreover, Cronbach's alpha improved if this item was deleted. However, the item was kept, since it was actually the same question posed in 2A, but applied to the subject of natural sciences, instead of mathematics. In addition, it should be noted that having only 4 items in this questionnaire may have contributed to the low Cronbach's alpha coefficient.

3.2.2. Analysis of Qualitative Information

The goal is to provide meaningful feedback about the respondents' thought processes when responding to survey items. Then, it is necessary to gather evidence that survey items and response options are well understood by respondents

TABLE 8 | Mean (teachers-post).

Media	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15
Representative	5.67	4.67	5.67	5.5	6	6	5.67	6	5.83	5.83	5.67	6	5.83	6	5.5
Relevance	5.67	3.83	5.67	5.67	6	6	5.83	6	6	5.83	5.83	6	5	6	5.5
Formulation	6	3.7	5.83	5.83	5	6	6	6	5.17	5.83	6	6	5.83	6	6

TABLE 9 | Design of the questionnaires (v2).

Questionnaire	Dimensions (item number)	Scale	Analysis type
Parents	Overall impact (1–3)	2 multiple choice	Qualitative
		1 dichotomous (with open-ended question)	Qualitative
	Impact on parents (4–7)	4 Likert (1–5 points)	Quantitative
	Satisfaction and project improvement (8–10)	1 Likert (1–5 points)	Quantitative
		2 Open-ended questions	Qualitative
Students-pre	STEM interests (1–2)	1 dichotomous (with open-ended question)	Qualitative
		1 Likert (1–5 points)	Quantitative
	Self-efficacy: perceived achievement (3)	1 Likert (1–5 points)	Quantitative
	Achievement in STEM subjects (4)	1 open-ended question	Qualitative
Students-post	Degree of participation (1–2)	2 open-ended questions	Qualitative
	Impact on students (3–6)	4 Likert (1–5 points)	Quantitative
	Satisfaction and project improvement (7–9)	1 Likert (1–5 points)	Quantitative
		2 open-ended questions	Qualitative
Teachers-pre	Motivation toward the project (1)	1 open-ended question	Qualitative
	Expectations (students) (2–4)	3 open-ended questions	Qualitative
	Expectations (teachers) (5)	1 open-ended question	Qualitative
Teachers-post	Degree of participation (1–3)	2 open-ended questions	Qualitative
		1 multiple choice answer	Qualitative
	Impact on students (4–6)	3 open-ended questions	Qualitative
	Impact on teachers (7–14)	1 open-ended question	Qualitative
		1 multiple choice answer	Qualitative
		6 Likert (1–5 points)	Quantitative
	Satisfaction and project improvement (15–17)	1 Likert (1–5 points)	Quantitative
		2 open-ended questions	Qualitative

The second column includes the dimensions and the item number in parentheses. The scale and the type of analysis are included in the third and fourth column, respectively.

Wolf et al. (2021). From the qualitative data, the answers given by all the participants were analyzed in parallel by each researcher to determine how the questionnaires worked in a real sample and to refine items if necessary. Researchers assessed the following questions for the items that had not been answered on a Likert scale in each questionnaire:

q1. If the item was understood and corresponded to the measured dimension. In this way, it is possible to have evidence of face validity i.e., to recognize the pertinence of the evaluation system by analyzing the answers given. The researchers indicated yes or no. In case of a negative answer, the reasons were noted down.

q2. If there were responses that could suggest presenting the item in another format or with some change in its presentation, in order to improve it. If they considered it appropriate, they suggested the reasons.

q3. Observations, if they considered any comment necessary, when they had answered "no" in any of the previous items.

Table 11 synthesizes by questionnaires and items the proposals of the group of 6 researchers. The columns "Relevance of the evaluation system" and "Presentation format" indicate the number of yes respondents from the 6 researchers. The last column, "comments," includes the observations when the researchers disagreed or any other comments they considered of interest.

In general terms, it can be seen that all the responses to the items building the questionnaires met the objective for which they were designed, since all six researchers agreed that, after analyzing all the results, there was no response that did not meet the indicator. They also agreed that the presentation format was adequate in most of the items, but some needed to be revised. Fifty percent of the researchers proposed to modify the

TABLE 10 | Summary of the Cronbach's alpha results in phase II.

Questionnaire	N of items	N valid / N samples (cases)	Cronbach's alpha	Items (questionnaire)	Corrected item (Total correlation)	Cronbach's alpha if item deleted
Parents	5	112 / 113	0.85	4	0.55	0.85
				5	0.73	0.80
				6	0.73	0.80
				7	0.67	0.81
				8	0.63	0.83
Students-pre (Primary)	4	32 / 32	0.49	2A	0.42	0.25
				2B	0.05	0.59
				3A	0.43	0.24
				3B	0.25	0.45
Students-pre (Secondary)	6	218 / 236	0.82	2A	0.56	0.79
				2B	0.52	0.80
				2C	0.66	0.77
				3A	0.55	0.79
				3B	0.56	0.79
				3C	0.61	0.78
Students-post	5	220 / 220	0.8	3	0.67	0.73
				4	0.65	0.74
				5	0.58	0.76
				6	0.36	0.82
				7	0.67	0.73
Teachers-post	6	14 / 14	0.65	9	0.33	0.63
				10	0.18	0.66
				11	-	-
				12	0.07	0.68
				13	0.71	0.46
				14	0.66	0.47
				15	0.49	0.61

type of response in three items: i) in the initial questionnaire for students, item 4 (performance in STEM subjects); ii) in the final questionnaire for students, item 2 (degree of participation); and iii) in the final questionnaire for teachers, item 2 (degree of participation). In addition, other comments were raised in item 1 and 2 of the overall impact on parents, since some of the multiple-choice answers were not chosen, as indicated in the table. Following the parallel analysis, the researchers participated in a debriefing until a consensus was reached on the changes needed. The results and conclusions of the discussion were as follows:

- **Parents questionnaire**. One of the researchers suggested that some of the multiple-choice options were not selected by any subject. Although she considered that the presentation format was adequate, she offered this topic for discussion. Researchers agreed that since there was a possibility that some person may point out these options in another sample, the presentation format should be maintained.
- **Students-pre questionnaire.** Fifty percent of the researchers suggested modifying the presentation format in the achievement in STEM subjects (item 4). In the discussion it became clear that it was a numerical response and that the open response option caused some students to indicate values with decimals, others in intervals, others suggested not remembering their grade and even subjective sentences such as "very bad grade". In the Spanish educational system, in secondary education, the optional nature of some subjects means that they are not prescriptive for all students. Therefore, in order to improve the coding and interpretation of the results, researchers agreed to present this item as a multiple-choice response with the following options: 0–3, 3.1–4.9, 5–5.9, 6–6.9, 7–8.9, 9–10, and *I do not take this course.*
- **Students-post and teachers-post questionnaires.** Item 2 measuring the degree of participation was discussed in both questionnaires. Fifty percent of the researchers suggested a closed response. Similar to the students-pre questionnaire discussion, a multiple-choice presentation format was decided, since it was seen that some answers provided intervals of hours of participation, or subjective sentences such as "many" or "the class hours." To avoid difficulties in processing the information, the multiple options were specified as follows: 0–1 h, between 1 and 2 h, between 2 and 3 h, between 3 and 4 h, between 4 and 5 h, between 5 and 6 h, between 6 and 7 h, between 7 and 8 h, between 8 and 10 h, between

TABLE 11 | Qualitative analysis (v2).

Questionnaire	Dimensions	Item	Relevance	Presentation format	Comments
Parents	Overall impact	1	6	6	No answer “nothing" or “other"
		2	6	6	No answer “other"
		3	6	6	
	Satisfaction and project improvement	9	6	6	
		10	6	6	
Students-pre	STEM interests	1	6	6	
	Achievement in STEM subjects	4	6	3	Modify to closed response (multiple choice)
Students-post	Degree of participation	1	6	6	
		2	6	3	Modify to closed response (multiple choice)
	Satisfaction and project improvement	8	6	6	
		9	6	6	
Teachers-pre	Motivation toward the project	1	6	6	
	Expectations (students)	2	6	6	
		3	6	6	
		4	6	6	
	Expectations (teachers)	5	6	6	
Teachers-post	Degree of participation	1	6	6	
		2	6	3	Modify to closed response (multiple choice)
		3	6	6	
	Impact on students	4	6	6	
		5	6	6	
		6	6	5	Add: “justify your answer" (some subjects indicate “positively" without explanation)
	Impact on teachers	7	6	6	
		8	6	6	
	Satisfaction and project improvement	16	6	6	
		17	6	6	

Dimensions and number of items are included in the second and third column, respectively. The fourth and fifth columns collect the number of positive answers in each dimension, relevance and presentation format, respectively. Observations raised by the researchers are included in the last column.

10 and 15 h and more than 15 h. These intervals were established based on the analysis of the answers given in the pilot sample. Finally, in the teachers-post questionnaire, a researcher suggested including “justify your answer" in item 6 on the impact on students, since she appreciated that some of the answers evaluated the project “positively" without providing arguments. The suggestion was accepted by the rest of the researchers, so the formulation of the question was modified.

4. DISCUSSION

The research presented in this paper aims at contributing to the state of the art of informal STEM education by describing the process of how to obtain evidences of reliability and validity of a set of instruments. This set of instruments comprises five questionnaires for the evaluation of the impact of the *family* action from the Girls4STEM initiative, which includes all the participants: students, families (parents) and teachers. The initial specific objectives of this research have been fulfilled. Firstly, in phase I, the initial version (v1) of the questionnaires has been designed, considering the initiative’s objectives and important dimensions to measure. The five questionnaires have been subjected to an expert judgment, to obtain evidence of validity of these instruments and to refine them if necessary. The results of all of them suggest high content validity through the calculation of the CVR, means and inter-rater reliability, which confirms the consistency of the results. Nevertheless, it has been necessary to delete some of the items, as well as to reformulate others. Specifically, the following changes have been necessary in the debugging process:

- Parents questionnaire: reformulation of items 2 and 5, given their means in the formulation dimension.
- Students-pre: reformulation of items 2 and 3, given their means in the formulation dimension. In addition, a new item on perceived achievement in STEM subjects has been added.
- Students-post: deletion of items 1 and 2, due to their CVRs values and their low means in representativeness and relevance. Two new items have been constructed from open-ended questions to determine the degree of participation (given that former items 1 and 2 were dealing with this metric).

The order of items 4 and 5 has been changed, following the proposal in the open-ended questions.

- Teachers-pre: deletion of item 2, due to its CVR, in addition to the fact that the means in representativeness and relevance pointed to a need for reformulation.
- Teachers-post: reformulation of item 2 due to its representativeness, relevance and formulation means. Former item 2 has been split into two new items.

Despite the modifications, all the questionnaires in version (v2) measure the dimensions proposed in **Table 2**, except the initial questionnaire for students, which includes a new dimension, the perception of competence (self-efficacy). In addition, there are some changes in the number of items, as the initial questionnaire for students goes from 3 to 4 items, the initial questionnaire for teachers reduces one item in (from 6 to 5) and the final questionnaire for teachers increases in one item (from 16 to 17). The design and feature of the questionnaires in version (v2) has been given in **Table 9**.

Once the objective of designing the instruments in phase I has been achieved and sufficient evidence of content validity has been obtained in this expert judgment, the analysis of the questionnaires in version (v2) has been carried out in a pilot sample. The pilot sample contains students from all pre-university academic cycles (primary, secondary), is gender balanced in line with the inclusive spirit of the project, and the schools are located in diverse contexts (from small urban centers to large cities).

The results regarding the evidence of reliability in the applied sample suggest that there is sufficient internal consistency of the Likert-type items included in each of the questionnaires. After the qualitative analysis of the remaining items, it is concluded that they have been answered in their entirety, in accordance with the purpose for which they were designed, so that the administration of the questionnaires to the pilot sample allows us to conclude that the objective of phase II has been achieved. In spite of this, it is necessary to modify some of the response formats. Specifically, in the initial student questionnaire, item 4 has changed from an open-ended question to a multiple-choice response to avoid the broad range of responses that has been observed when processing the qualitative analysis. The same happens with item 2 of the final questionnaire for students and teachers. In addition, item 6 of the teachers-post questionnaire adds the suggestion "justify your answer" to improve the quality of the gathered data. As a result of phase II, the version (v3) of the five questionnaires has been obtained, where the students-pre questionnaire, and the teachers-pre and teachers-post questionnaires have been modified as discussed above with respect to version (v2).

The set of questionnaires, in their final version (v3), are a valuable resource for the evaluation of the *family* action of the Girls4STEM initiative, allowing to assess the impact over all target audiences (students, families and teachers). The mixed methods methodology has allowed to refine the set of instruments through the use of different techniques, such as the expert judgment. Moreover, the analysis of the set of instruments administered to a pilot sample of the study population has enabled the collection of evidence that survey items and response options are well understood by respondents.

This set of instruments has been designed and validated with the aim of overcoming the challenges faced by the evaluation of informal STEM education actions. On the one hand, the instruments incorporate features in the evaluation that are often overlooked, such as improvement of the initiative, with measures at different times, e.g., pre and post action for students and teachers. On the other hand, completing the questionnaires does not require excessive time due to their well-designed formulation, which maximizes the likelihood that they will be completed properly by the participants, including primary students from lower courses which might be less familiar with filling on-line forms without help. The fact that they can be delivered on-line, simplifies the posterior data analysis and contributes to the sustainability of the initiative. In addition, preliminary reliability and validity evidence conducted by a multidisciplinary team of researchers has been provided, which to the best of our knowledge, positions this work as a core reference in informal STEM education contexts. Although the initiative Girls4STEM is located in Spain, the process followed to achieved the set of instruments in version (v3) can be applied to any informal evaluation initiative with a low-cost implementation. Moreover, the set of instruments is openly offered for review or administration in other educational experiences in informal education, so that particular features of different cultural contexts can be incorporated via each initiative's objectives. Nevertheless, it is desirable to continue researching and collecting new evidence in on-going and future editions of the initiative, in order to continue improving the rigor of the questionnaires, being applied to other samples or adapted for administration to other STEM educational projects.

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 authors.

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 from the patients/ participants or patients/participants legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MH-P designed the research. CB-M, EV, and XB collected the data. MH-P, CB-M, EV, EL-I, AF, XB, and SR analyzed the research. MH-P, CB-M, and EL-I searched the literature.

MH-P, CB-M, EV, and EL-I wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partially supported by the project FCT-20-15904 from the Fundación Española para la Ciencia y la Tecnología (FECYT) and the Ministerio de Ciencia e Innovación and the project GV/2021/110 from Generalitat Valenciana.

ACKNOWLEDGMENTS

The authors would like to thank all the entities that support the Girls4STEM initiative: the Vice-Principal of Equality, Diversity and Sustainability (University of Valencia); the Equality Unit (University of Valencia); the Scientific Culture and Innovation Unit (University of Valencia); and the Center for Training, Innovation, and Educational Resources in the Scientific, Technological, and Mathematical fields (CEFIRE STEM, Conselleria of Education, Research, Culture, and Sports of the Generalitat Valenciana), as well as the School of Engineering from the University of Valencia. Special thanks to all the primary and secondary schools that have participated in the three editions, the STEM experts, the project's sponsors, and last but not least, all the colleagues working in the project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.937058/full#supplementary-material

REFERENCES

AERA (2014). *Standards for Educational and Psychological Testing*. Lanham, MD: American Educational Research Association.

Allen, S., and Peterman, K. (2019). Evaluating informal STEM education: Issues and challenges in context. *New Direct. Evaluat*. 2019, 17–33. doi: 10.1002/ev.20354

Anguera, M. T., Camerino, O., and Castañer, M. (2012). "Mixed methods procedures and designs for research on sport, physical education and dance," in *Mixed Methods Research in the Movement Sciences: Case Studies in Sport, Physical Education and Dance* (London), 3–27.

Ayuso, N., Fillola, E., Masi,á, B., Murillo, A. C., Trillo-Lado, R., Baldassarri, S., et al. (2021). Gender gap in STEM: a cross-sectional study of primary school students' self-perception and test anxiety in mathematics. *IEEE Trans. Educ*. 64, 40–49. doi: 10.1109/TE.2020.3004075

Benavent, X., de Ves, E., Forte, A., Botella-Mascarell, C., López-Iñesta, E., Rueda, S., et al. (2020). Girls4STEM: Gender diversity in STEM for a sustainable future. *Sustainability* 12, 6051. doi: 10.3390/su12156051

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

Botella, C., López-Iñesta, E., Rueda, S., Forte, A., de Ves, E., Benavent, X., et al. (2020). "Iniciativas contra la brecha de género en STEM," in *Una guía de buenas prácticas*. [*Initiatives against the gender gap in STEM. A guide to good practices*]. *In XXVI Jornadas sobre la Enseñanza Universitaria de la Informática (JENUI)* (Valencia).

Botella, C., Rueda, S., López-Iñesta, E., and Marzal, P. (2019). Gender diversity in STEM disciplines: a multiple factor problem. *Entropy* 21, 30. doi: 10.3390/e21010030

Botella-Mascarell, C., López-Iñesta, E., Rueda, S., Forte, A., de Ves, E., and Benavent, X. (2021). "Using databases to improve the visibility of women's contributions in STEM fields," in *8th ACM Celebration of Women in Computing: womENcourage*. Available online at: https://roderic.uv.es/handle/10550/80281.

Colás Bravo, P., and Villaciervos Moreno, P. (2007). La interiorización de los estereotipos de género en jóvenes y adolescentes. [The internalization of gender stereotypes in young people and adolescents]. *Revista de Investigación Educativa* 25, 35–38. Available online at: https://revistas.um.es/rie/article/view/96421

Creswell, J. W., and Clark, V. L. P. (2003). *Designing and Conducting Mixed Methods Research, 2nd Edn*. Thousand Oaks, CA: Sage Publications.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334. doi: 10.1007/BF02310555

Cronbach, L. J., and Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educ. Psychol. Meas*. 64, 391–418. doi: 10.1177/0013164404266386

Diekman, A. B., Clark, E. K., and Belanger, A. L. (2019). Finding common ground: Synthesizing divergent theoretical views to promote women's STEM pursuits. *Soc. Issues Policy Rev*. 13, 182–210. doi: 10.1111/sipr.12052

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, 142–175. doi: 10.1177/1088868316642141

Fàbregues, S., Paré, M.-H., and Meneses, J. (2019). Operationalizing and conceptualizing quality in mixed methods research: a multiple case study of the disciplines of education, nursing, psychology, and sociology. *J. Mix Methods Res*. 13, 424–445. doi: 10.1177/1558689817751774

Fu, A. C., Kannan, A., Shavelson, R. J., Peterson, L., and Kurpius, A. (2016). Room for rigor: Designs and methods in informal science education evaluation. *Visitor Stud*. 19, 12–38. doi: 10.1080/10645578.2016.1144025

George, D., and Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Study Guide and Reference, 17.0 Update, 10th Edn*. Boston, MA: Pearson Education.

Gladstone, J. R., and Cimpian, A. (2021). Which role models are effective for which students? a systematic review and four recommendations for maximizing the effectiveness of role models in STEM. *Int. J. STEM Educ*. 8, 10127–10134. doi: 10.1186/s40594-021-00315-x

González-Pérez, S., Mateos de Cabo, R., and Sáinz, M. (2020). Girls in STEM: Is it a female role-model thing? *Front. Psychol*. 11, 2204. doi: 10.3389/fpsyg.2020.02204

Grack Nelson, A., Goeke, M., Auster, R., Peterman, K., and Lussenhop, A. (2019). Shared measures for evaluating common outcomes of informal STEM education experiences. *New Direct. Evaluat*. 2019, 59–86. doi: 10.1002/ev.20353

Griffiths, A. J., Brady, J., Riley, N., Alsip, J., Trine, V., and Gomez, L. (2021). STEM for everyone: A mixed methods approach to the conception and implementation of an evaluation process for STEM education programs for students with disabilities. *Front. Educ*. 5, 545701. doi: 10.3389/feduc.2020.545701

Groenvold, M., Klee, M. C., Sprangers, M. A., and Aaronson, N. K. (1997). Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. *J. Clin. Epidemiol*. 50, 441–450. doi: 10.1016/S0895-4356(96)00428-3

Guenaga, M., Eguíluz, A., Garaizar, P., and Mimenza, A. (2022). The impact of female role models leading a group mentoring program to promote STEM vocations among young girls. *Sustainability* 14, 1420. doi: 10.3390/su14031420

Gwet, K. L. (2014). *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters*. Gaithersburg: Advanced Analytics, LLC.

Habig, B. (2020). Practical rubrics for informal science education studies: (1) a STEM research design rubric for assessing study design and a (2) STEM impact rubric for measuring evidence of impact. *Front. Educ*. 5, 554806. doi: 10.3389/feduc.2020.554806

Hargraves, R. H., Morgan, K. L., Jackson, H., Feltault, K., Crenshaw, J., McDonald, K. G., et al. (2021). What's your STEMspiration?: adaptation and validation of a survey instrument. *Front. Educ.* 6, 667616. doi: 10.3389/feduc.2021.667616

Haynes, S. N., Richard, D., and Kubany, E. S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychol. Assess.* 7, 238. doi: 10.1037/1040-3590.7.3.238

Herce Palomares, M. P., and Román González, M. (2021). "Educación del talento STEM (ciencia, tecnología, ingeniería y matemáticas) como impulso de los Objetivos de Desarrollo Sostenible (ODS)," in [*STEM (Science, Technology, Engineering and Mathematics) Talent Education as a Driver of the Sustainable Development Goals (SDGs)*] (Cham: Dykinson).

Herce Palomares, M. P., Román González, M., and Giménez Fernández, C. (2022). El talento STEM en la educación obligatoria: una revisión sistemática. [STEM talent in k-10: a systematic review]. *Revista de Educ.* 396, 537-549. doi: 10.4438/1988-592X-RE-2022-396-530

Kier, M. W., Blanchard, M. R., Osborne, J. W., and Albert, J. L. (2014). The development of the STEM career interest survey (STEM-CIS). *Res. Sci. Educ.* 44, 461–481. doi: 10.1007/s11165-013-9389-3

Lawshe, C. H. (1975). A quantitative approach to content validity1. *Pers. Psychol.* 28, 563–575. doi: 10.1111/j.1744-6570.1975.tb01393.x

Loewenthal, K., and Lewis, C. (2001). *An Introduction to Psychological Tests and Scales, 2nd Edn*. London: Psychology Press.

López-Iñesta, E., Botella, C., Rueda, S., Forte, A., and Marzal, P. (2020). Towards breaking the gender gap in science, technology, engineering and mathematics. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje* 15, 233–241. doi: 10.1109/RITA.2020.3008114

López-Iñesta, E., Queiruga-Dios, M. Á., García-Costa, D., and Grimaldo, F. (2022). Citizen science projects: An opportunity for education in scientific literacy and sustainability. *Mètode Sci. Stud. J.* 12, 32–39. doi: 10.7203/metode.12.17824

National Research Council (2009). *Learning Science in Informal Environments: People, Places, and Pursuits*. Washington, DC: The National Academies Press.

Onwuegbuzie, A. J., and Johnson, R. B. (2006). The validity issues in mixed research. *Res. Sch.* 13, 48–63.

Robinson, A., and Murray, N. (2019). *Evaluating Ocean Learning–The Principles and Practicalities of Evaluating Formal Education Audiences in an Informal Education Environment*. Cham: Springer International Publishing.

Sáinz, M., Martínez-Cantos, J.-L., Rodó-de Zárate, M., Romano, M. J., Arroyo, L., and Fàbregues, S. (2019). Young spanish people's gendered representations of people working in STEM. A qualitative study. *Front. Psychol.* 10, 996. doi: 10.3389/fpsyg.2019.00996

Sáinz, M., and Müller, J. (2018). Gender and family influences on spanish students' aspirations and values in STEM fields. *Int. J. Sci. Educ.* 40, 188–203. doi: 10.1080/09500693.2017.1405464

Santos, E. D. D., Albahari, A., Díaz, S., and Freitas, E. C. D. (2021). 'Science and technology as feminine': raising awareness about and reducing the gender gap in STEM careers. *J. Gender Stud.* 31, 1–14. doi: 10.1080/09589236.2021.1922272

Schwarzer, R., and Baessler, J. (1996). Evaluación de la autoeficacia: adaptación española de la escala de autoeficacia general. [Evaluation of self-efficacy: Spanish adaptation of the general self-efficacy scale]. *Ansiedad Estrés* 2, 1–8.

Shekhar, P., Prince, M., Finelli, C., Demonbrun, M., and Waters, C. (2019). Integrating quantitative and qualitative research methods to examine student resistance to active learning. *Eur. J. Eng. Educ.* 44, 6–18. doi: 10.1080/03043797.2018.1438988

Skjong, R., and Wentworth, B. (2000). "Expert judgement and risk perception," in *Offshore and Polar Engineering Conference*, eds R. Skjong and B. Wentworth (Seattle, WA: ISOPE), 537–544.

Tena Gallego, È., and Couso, D. (2019). *Compendio de herramientas para evaluar el impacto de iniciativas de investigación científica para alumnado de educación primaria con perspectiva RRI. [Compendium of tools for assessing the impact of scientific research initiatives for primary school students with a RRI perspective]*. Working document. Available online at: https://ddd.uab.cat/record/210969?ln=es

Tristán-López, A. (2008). Modificación al modelo de Lawshe para el dictamen cuantitativo de la validez de contenido de un instrumento objetivo. [Modification to Lawshe's model for quantitative judgment of the content validity of an objective instrument.]. *Avances en Medición* 6, 37–48.

Trizano-Hermosilla, I., and Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Front. Psychol.* 7, 769. doi: 10.3389/fpsyg.2016.00769

UNESCO (2017). *Cracking the code: Girls' and womens' education in Science, Technology, Engineering and Mathematics (STEM)*. Available online at: http://unesdoc.unesco.org/images/0025/002534/253479e.pdf (accessed May 5, 2022).

Wolf, M. G., Ihm, E., Maul, A., and Taves, A. (2021). "Survey item validation," in *The Routledge Handbook of ResearchMethods in the Study of Religion*, 2nd Edn, eds S. Engler and M. Stausberg (London: Routledge).

Xiao, L., and Hau, K.-T. (2022). Performance of coefficient alpha and its alternatives: effects of different types of non-normality. *Educ. Psychol. Meas.* 1–23. doi: 10.1177/00131644221088240

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

TYPE Hypothesis and Theory
PUBLISHED 12 August 2022
DOI 10.3389/fpsyg.2022.918439

OPEN ACCESS

EDITED BY
Sergi Fàbregues,
Open University of Catalonia, Spain

REVIEWED BY
Kathryn Holmes,
Western Sydney University,
Australia
Katja Upadyaya,
University of Helsinki,
Finland

*CORRESPONDENCE
Susana González-Pérez
susana.gonzalezperez@ceu.es

SPECIALTY SECTION
This article was submitted to
Gender, Sex and Sexualities,
a section of the journal
Frontiers in Psychology

RECEIVED 12 April 2022
ACCEPTED 07 July 2022
PUBLISHED 12 August 2022

CITATION
González-Pérez S, Martínez-Martínez M,
Rey-Paredes V and Cifre E (2022) I
am done with this! Women dropping out of
engineering majors.
Front. Psychol. 13:918439.
doi: 10.3389/fpsyg.2022.918439

I am done with this! Women dropping out of engineering majors

Susana González-Pérez[1]*, Miryam Martínez-Martínez[1], Virginia Rey-Paredes[2] and Eva Cifre[3]

[1]Department of Business Economics, School of Economics and Business, Universidad San Pablo-CEU, CEU Universities, Madrid, Spain, [2]Department of Economics, School of Economics and Business, Universidad San Pablo-CEU, CEU Universities, Madrid, Spain, [3]Department of Developmental, Educational, Social Psychology and Methodology, Faculty of Health Science, Universitat Jaume I, Castellón, Spain

Women are still underrepresented in STEM careers (Science, Technology, Engineering, and Mathematics). One of the possible drivers behind this gender gap in the labour market is the female dropout from STEM education. The causes of the gender differences in the persistence of pursuing STEM studies have been explained by multiple factors related to interest and resolution in this type of career. The goal of the present research is to study the Engineering persistence gender gap in higher education by exploring the main factors underlying the leakage in the pipeline of Engineering fields. Our study reports the results of 34 qualitative in-depth interviews where internal barriers, stereotypes and external obstacles are assessed by women who have left their university degrees, compared with men who have withdrawn and women who have persisted. Results from the content analysis suggest that the undermining of persistence in Engineering fields is related to factors such as the chilly and hostile environment in classes or the workload from an excessively demanding curriculum. Other factors affecting women's withdrawal are the lack of role models and the perceived incongruity between the female gender role and STEM roles in society, leading to a weakening of female students' self-efficacy and eroding their sense of belongingness, even making them consider dropping out of their Engineering degree. These findings provide information for the design of future STEM interventions aimed to enhance women's persistence in STEM university studies.

KEYWORDS

STEM, self-efficacy, belongingness, motivation, curriculum, gender stereotypes, persistence

Introduction

The need for Science, Technology, Engineering, and Mathematics (STEM) development to increase and maintain our current quality of life is globally acknowledged. STEM fields are the basis of our everyday lives, being responsible, for instance, for having clean water, food to eat and life-saving medicines. However, most of these advances have historically been seen as male domains, with a clear need to extend them to women. According to Eurostat (2021), in 2019, out of a total of 15.4 million posts in Science & Engineering (including physicists, mathematicians, life science professionals, and engineers), there were only 6.3 million female scientists and engineers accounting for 41% of all employment in the European Union. In the OECD countries, the percentage of graduates in Engineering reaches 14%, however, the composition is very different between men and women. Among men, the percentage reaches 25%, while among women only 7%. A situation is very similar to that found in Spain, where women in engineering are 6% compared to 22% of men. However, there are OECD countries where this situation has been mitigated, as is the case of Iceland (OECD, 2019). STEM fields are not only essential to improve our quality of life; they are expected to grow by 10.5% between 2020 and 2030 in the United States (US Bureau of Labor Statistics, 2021). Currently, the top 25 college degrees by pay and demand are all in STEM subjects (World Economic Forum, 2021), mainly in male-dominated jobs (i.e., top 5: Architectural Engineering, Construction Services, Computer Engineering, Aerospace Engineering, and Transportation Sciences and Technologies). Despite this, a large number of STEM undergraduates drop out of their fields. Compared with non-STEM studies, STEM drop-out is in general higher; this is even more marked in the case of women, with female students displaying a 23% higher drop-out rate than their male counterparts, even though these female students in STEM appear to be positively selected in terms of study capital (Isphording and Qendrai, 2019). Barriers to entry into STEM education for female students have gained strong attention from education researchers, such as gender stereotypes (i.e., Master et al., 2016), prior achievement and attitudes (Marsh et al., 2019), or in general, social factors, institutional structures, poor advice and early education environments (Blackburn, 2017). However, subsequent gender differences in continuing to pursue STEM studies are less well studied. Even interventions performed to improve graduation rates for students in STEM (i.e., Hamm et al., 2020) do not consider gender a core issue (just as a control variable). Among all the STEM disciplines, Engineering is one of the most male-dominated due to its well-recognized difficulty (Center of Research and Educational Documentation and Women's Institute, 1988; López-Sáez et al., 2011). Research in higher education indicates that there are many factors that influence female students' retention in Engineering (Blickenstaff, 2005; Eddy and Brownell, 2016), contributing to perpetuating the gender gap in this type of major. The lack of women in Engineering is a concern shared by public and educational institutions (Chavatzia, 2017) Despite such increasing efforts to find solutions, we continue to struggle to understand the reasons that lead women to leave Engineering degrees (Beasley and Fischer, 2012; Makarem and Wang, 2020). This study aims to tackle the gender gap in progression in Engineering majors, drawing on two theoretical approaches from both motivation and gender studies through a mixed qualitative-quantitative methodology that will allow delving into the Engineering drop-out phenomenon for female students. Thus, this study aims to address this gap by examining the reasons why female students withdraw from Engineering studies, and to find effective interventions to improve retention rates.

Summing up, this paper contributes to the literature through three main points:

1. Analyzing gender differences in drop out from Engineering majors, but also comparing with female students who continue their studies.
2. Giving a voice to female students who have dropped out of such careers, using the in-depth interview technique to listen to their personal experiences.
3. Providing outcomes that will facilitate the design of adjusted interventions. Understanding the reasons for dropping out will help to devise more effective and efficient strategies for women to stay in STEM careers.

Engineering dropout: Student motivation and the Tinto's persistence model

Students do not seek to be retained, they seek to persist (Tinto, 2015). Students' persistence could not happen without proper motivation, as Engineering students are faced with a very challenging curriculum (Wang and Degol, 2013; Kelly, 2016) and need to expend a huge effort to succeed in their studies. To increase completion rates, institutions are required to adopt a student perspective, to understand how students' motivation is shaped and which measures or interventions can be addressed to enhance this motivation. When adopting this student perspective, they must do so taking into account that the reality of women and men is not the same, i.e., adopting a gender perspective. According to the European Institute for Gender Equality (EIGE) (n.d.), this means taking gender-based differences into account when looking at any social phenomenon, policy, or process. We will base our study on two different but complementary theories: one that comes from the field of education (Tinto, 1975, 2015) and another from gender studies (social role theory, Eagly, 1987a) to explain and promote the persistence of female students in Engineering majors. Tinto's model is a conceptually useful framework to analyze student dropout since it reflects the process that an undergraduate student experiences between the decision to abandon or continue their studies. This scheme is based on the fact that in order to continue and be successful, students must integrate socially and academically into the university. Thus, Tinto integrates the academic and social perspectives and ties them together to explain student dropout. The social role theory argues that the proximal causes of sex differences in individual behavior are framed by gender roles or the shared beliefs that apply to individuals based on their socially identified sex (Eagly, 1987a,b; Eagly et al., 2004). Because in all cultures women and men tend to specialize in different behaviors, people have different beliefs about what each sex can and should do, i.e., gender roles are descriptive and prescriptive (Wood and Eagly, 2010). These beliefs constitute socially shared stereotypes within a society, meaning that gender roles are reflected in a society's stereotypes about men and women. As Wood and Eagly (2010) point out, the descriptive aspect of gender roles indicates what is typical for each sex, so people rely on this information when they are concerned about what is normal for their sex. On the other side, the prescriptive aspect describes what is desirable and admirable for ache sex, so people rely on this information when are motivated to gain social approval or to bolster their own esteem. Thus, in the case of Engineering, women might feel that they are not accomplishing their gender social role of nurturing and they do not identify that Engineering may lead to fulfilling communal goals. In this way, social role theory in general, and the (communal) goal congruity perspective (Diekman and Steinberg, 2013) in particular, will work as a gender mainstreaming theory while developing research questions related to Tinto's persistence theory to understand how women's entry, engage and exit of a specific social role (i.e., STEM career).

To be motivated means to be moved to do something (Ryan and Deci, 2000). It is a factor that leads behavior and determines its direction, force, and insistence (Sevinc et al., 2011). Also, motivation is a theoretical concept that is used to explain the beginning, direction, force, and insistence of goal-oriented behavior (Brophy, 2004). This insistence on goal-oriented behavior or persistence is a manifestation of motivation (Bandura, 1989). However, whereas early experiences and goals can lead students to choose Engineering, this motivation changes over time, as students face different college experiences that may affect their willingness to persist. Tinto (2015) proposed that persistence is driven by motivation, which is determined by the lower-order factors of self-efficacy perception, sense of belonging, perceived worth, and relevance of the curriculum. At the heart of Tinto's model, there is the idea that to be successful, and therefore persist, a student must be well integrated both socially and academically into the college system (Tinto, 1975, 1987). Nevertheless, this integration does not come only from the student but also from the educational environment. In this sense, the social and academic factors linked to the educational environment can help or hinder the integration of students (Casad et al., 2019). This is a useful framework within which not only to investigate the process of student attrition and persistence in Engineering but also to identify possible interventions to reduce withdrawal. Tinto's perspective requires a holistic approach to studying dropouts taking in different kinds of factors, highlighting that this withdrawal is a process in which it is possible to act (Tinto, 1998). Tinto's model remains one of the most widely used and cited models in understanding and explaining students' dropouts (Braxton and Hirschy, 2004; Keup, 2005; Bensimon, 2007; D'Amico et al., 2014), also regarding those in STEM disciplines (i.e., Nicoletti, 2019; Johnson et al., 2020). In particular, Botanga et al. (2021) found in a sample of underrepresented students minority (URM) of STEM that Tinto's model focuses on integration and belonging, but fails to theorize concepts related to student agency, racial identity, and racism, so important in this URM group. These results stress the need to adapt the model to different samples, such as we do in the case of women in STEM.

Engineering dropout: The persistence model from a gender perspective

The social role theory (Eagly, 1987a) offers a framework to understand how features of social roles intersect with individuals' goals and role pursuits in general, and of women in STEM in particular. Therefore, our research adapts Tinto's model of student motivation and persistence (Tinto, 2015) considering gender role congruity, so, student's circumstances and inputs should be included, incorporating their perspective on the main barriers and obstacles that might have undermined their motivation and their willingness to persist in their academic trajectories from a gender perspective (Figure 1).

Tinto (1987) posited that a departure decision was a process by which a student who experienced lack of motivation decided not to persist with college studies. One of the first and more important input variables to determine students' motivation are goals. According to Locke and Latham's (1990) of goal setting theory, a goal is defined as what the individual is consciously trying to do. Goals direct attention and action. Furthermore, challenging goals mobilize energy and lead to higher effort, motivating people to develop strategies that will enable them to perform at the required goal levels. Also, the goal accomplishment can lead to satisfaction and further motivation whereas their non-accomplishment can lead to frustration and lower motivation. Goals are shaped by early experiences (Tinto, 1975) as they help specify the orientations the student brings into the college. Goals are a very complex issue and should be approached as a multidimensional variable. Tinto (1998) identified family background, prior schooling and skills and abilities as the main factors shaping students' goals. In this line, the goal congruity perspective (Diekman and Steinberg, 2013; Diekman et al., 2017) may help to explain why women enter, engage in and exit STEM pursuits, as it provides a framework to understand how motives influence social role selection, and in turn how these social roles afford or impede the pursuit of goals. This perspective is based on the social role theory, which posits that sex differences in individual behaviour are framed by gender roles, or the shared beliefs that apply to individuals on the basis of their socially identified sex (Eagly, 1987a; Eagly et al., 2004). For both sexes, good fit to the opportunities afforded by their society yields rewards in terms of ease of completing important tasks and building satisfying interpersonal relationships, so individuals thus (consciously or not) are more likely to seek and attain the goals that are afforded by their roles (Diekman and Eagly, 2008). According to the goal congruity framework, an important aspect of STEM decisions is the belief that STEM careers do not fulfil communal, other-oriented goals (Diekman and Steinberg, 2013), which is not aligned (goal incongruity) with women's roles. So,

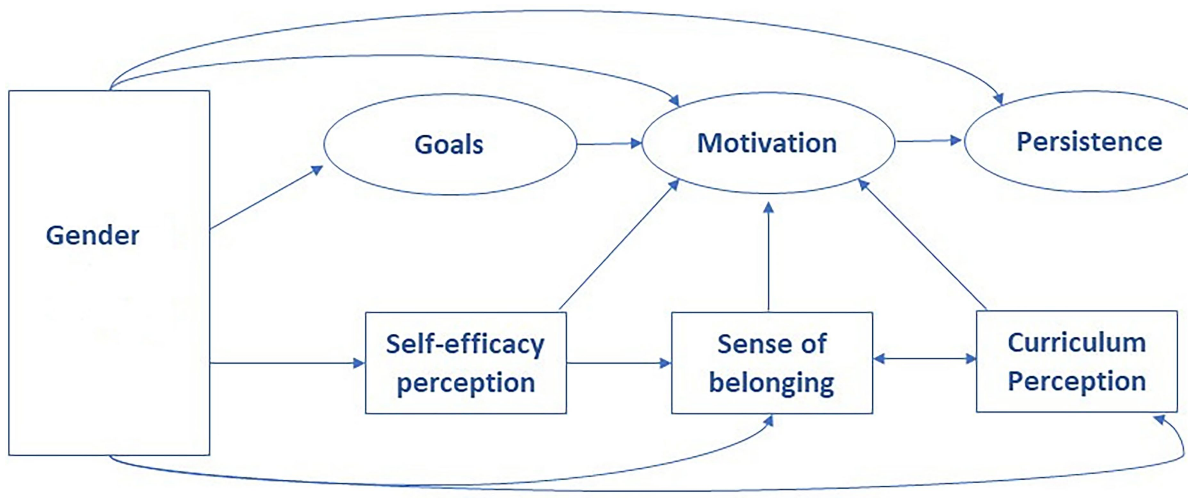

FIGURE 1
The proposed theoretical model of gendered role congruity in students' motivation and persistence in engineering.

these internalized values tend to drive female students away from male-stereotypic careers perceived lower in communion (Diekman et al., 2010), leading to horizontal segregation (Eagly, 1987b; Wood and Eagly, 2012). Also, as proposed by Tinto's model, the (communal) goal congruity considers the temporal dimension in three phases: (1) anticipate (in)congruity prior to role decisions, (2) experienced (in)congruity in a particular role, and (3) psychological and behavioral responses to maintain or seek congruity (Diekman et al., 2017). So, the socialization process plays a key role, as societal gender stereotypes (Bakan, 1966) lead even young females (Block et al., 2018) to internalize communal values instead of agentic traits (Eagly, 1987b) which are not congruent with the expected agentic traits associated to STEM. This anticipated goal incongruity may fuel the decision not to prior enroll in a STEAM career. The second phase focuses on what happens after individuals enter into social roles, i.e., STEM majors. Beliefs about anticipated goal (in)congruity might be more or less accurate with their actual experiences of goal (in)congruity. Then, in phase 3, individuals respond to maintain/seek congruity. To do so, they might change the motives (i.e., downplaying the importance of the motive, such as communality) or roles (i.e., dropping out STEM majors).

Motivation: Self-efficacy perception perspective

Among the other lower-order factors in Tinto's model that determine motivation, self-efficacy is especially important. Self-efficacy is an aspect of social cognitive theory defined as "the exercise of human agency through people's beliefs in their capabilities to produce desired effects by their actions" (Bandura, 1997) or "judgments of how well one can execute courses of action required to deal with prospective situations" (Bandura, 1982).

In this line, prior schooling and preferences for school subjects strongly influence whether women feel motivated to study STEM at college (Delaney and Devereux, 2019) based on their experience. Self-efficacy has been shown to mediate perseverance, as students who have higher self-efficacy are more likely to persist in the face of difficulty (Seymour and Hewitt, 1997; Zimmerman, 2000; Usher and Pajares, 2008). Regarding the sources of self-efficacy, this could be built by mastery experience in the classroom, i.e., by succeeding in a perceived to be very difficult exam or assignment (Usher and Pajares, 2008). Thus, the role of school subjects in later STEM enrolment contributes to the gender gap in STEM in college, making it critical to provide positive STEM experiences in school to increase female students' interest in STEM fields (Fervers et al., 2020). Skills and abilities also shape female students' goals, motivating them to enroll in a STEM major, for example, spatial skills (Halpern et al., 2007) and perceptions of ability have been found to predict career choices in Engineering (Eccles and Wigfield, 2002). Emotional or physiological states are also sources of self-efficacy that students may feel when completing a difficult task successfully (Phan, 2012; Phan and Ngu, 2016). Social persuasion (Bandura, 1997) is the external encouragement received from peers or faculty members. Finally, vicarious experience frequently occurs for students when one compares oneself with another peer (Bandura, 1997). In this line, sociologists have identified the so-called occupational inheritance phenomenon (Mannon and Schreuders, 2007), which shows that female students entering Engineering are more likely than men to have an engineer in the family, while those without engineers in the family must find another figure to inspire and motivate them. In sum, family, especially parents, are critical early socializers of their children's academic interests and their academic choices of STEM majors (Simpkins et al., 2006). Nevertheless, having the possibility to interact with role models (such as an engineer in the family or an inspiring science teacher) reduces the effect of gender stereotypes and increases intentions of female students' enrolment in STEM majors (González-Pérez et al., 2020). Also, intervention programs focused on self-efficacy sources have shown to be greatly successful to increase interest towards STEAM (STEAM + Arts) in female students (Ofori-Boadu, 2018).

Motivation: Sense of belongingness perspective

The third construct of Tinto's model, belongingness, has been linked with persistence at university (Webb et al., 2017). It is understood to be the sense of connectedness an individual experience within the learning environment (Osterman, 2000). In other words, belonging refers to students' sense of being accepted, valued, included, and encouraged by both teachers and peers and of feeling an important part of the group (Goodenow, 1993). The literature suggests that there is a strong relationship between the concept of belongingness and students' self-efficacy perception (St-Amand et al., 2017). The quality of belongingness is dependent on a variety of factors such as the level of a student's involvement in the different activities provided by college and the availability of support, which finally builds a sense of connection (Picciano, 2002). If these female students feel connected it will be more likely that they will be receptive and more deeply engage in learning (Roxburgh, 2012), while if female students experience greater uncertainty and feelings of not belonging, unsure of their social bonds and sensitive to rejection, they are at more risk of dropping out (Walton et al., 2015). There has been much research around the structural barriers that women face in STEM fields and that make them feel like 'strangers' in science (Massachusetts Institute of Technology, 1999; Sonnert et al., 2007).

Classroom climate affects students' belongingness perception and therefore their motivation and persistence in STEM (Shapiro and Sax, 2011). If female students in STEM face a difficult climate in the classroom, it could disproportionately affect them, producing feelings of depression and lowering self-confidence (Strenta et al., 1994). Women entering a male-dominated field may need to face social marginalization and may experience a climate in which they may feel unwelcome (Flam, 1991). Colbeck et al. (2001) suggest that this "chilly climate" for women in Engineering results from peer interactions, which is especially relevant as peer acceptance is a central concern in adolescence (Eaton et al., 1991). While it is perceived that peer interactions affect students more than faculty interactions, teachers also have an important role in making students feel comfortable and accepted in college, as well as promoting students' interest in their subjects (Astin and Sax, 1996).

Furthermore, women are more likely to leave STEM majors compared to men, in part because they lack similar role models such as teaching assistants and instructors (Marx and Roman, 2002). Women exhibit a self-perception of belonging in STEM culture and are more motivated to pursue studies in the presence of female role models (Stout et al., 2011). Nevertheless, men comprise the majority of STEM faculties that may not only signal that women do not belong or cannot succeed in these fields (Walton and Cohen, 2007) but also gives female students limited access to female faculty role models. In this context, with a limited representation of female students, even highly skilled and motivated women may wonder whether they belong on STEM major programs (Cheryan et al., 2009), which elicits that a more inclusive Engineering community is a crucial element so that female students do not feel alone (Ayre et al., 2013). However, faculty can also be a threat to female persistence in STEM fields, because as Hall and Sandler (1982) explain, faculty interactions can dampen women's ambitions, especially in male-dominated fields such as Engineering. Furthermore, researchers such as Wasburn and Miller (2004, 2005) have found that faculties treated male and female students differently, as they tend to be more condescending, and less respectful to female students. They have also found that faculties tend to exclude women from certain activities, for example giving them menial tasks within group projects. Grading criteria are often found unfair and biased, especially as female students feel they need to excel more than their male counterparts (Seymour and Hewitt, 1997). These authors found as well that ignoring, or tolerating misogynism, feel female students unwelcome in class. In both cases, both male-peer and faculty interaction might be biased due again to the goal and trait incongruence that they perceive between the female (communal) traits and what is thought to be successful for a masculinized career (agentic), which might in turn lead to prejudice against women (see for instance Eagly and Karau, 2002, about women leaders, i.e., in a masculinized career). Consequently, women might divert from STEM pathways because of gender stereotypes and prejudice (Diekman et al., 2017). For instance, both male and female science faculty have shown gender bias in preferring male over female applicants for a lab manager position, even when qualifications were matched experimentally (Moss-Racusin et al., 2012).

Motivation: Curriculum perception perspective

Curriculum, the last element of Tinto's model explaining motivation, is explained in turn and affected by the sense of belonging (Webb et al., 2017). Curriculum needs to be understood in an extended way, considering not only what is being taught (Kelly, 2009), but also the methods of assessment. Notably, students in STEM fields often earn lower grades than students in other fields (Ma and Liu, 2015). This is another factor that hindered female students studying for a degree in Engineering, as perception of low grades seems to be more discouraging for them than for their male counterparts (Lent et al., 2002). Moreover, the usage of curve-grading assignments encourages competition, where male students tend to feel significantly more comfortable, preventing cooperation and peer support (Guzdial et al., 2001).

Griffith (2010) found that students feel that classes in Engineering tend to be boring and needlessly difficult, forcing them to spend many hours and make huge efforts, sometimes without the desired outcome. This situation affects female students to a greater extent, undermining their perception of belonging and leading to them opting to drop out to a greater extent than their male counterparts (Seymour, 1995). Seron et al. (2018) discovered that during the first 2 years of Engineering majors, it would be necessary to incorporate as many real, everyday examples, as possible to continue to encourage women students to stick with challenging introductory classes (Wang and Degol, 2013; Kelly, 2016). In terms of the type of teaching methods, hands-on projects are more meaningful and interesting for female students (Mitchell, 1993; Halpern, 2004; Geist and King, 2008). Project-based learning (Blumenfeld et al., 2000) has proven to increase students' engagement as well as a deeper understanding of scientific problems (Kaldaras et al., 2021). Instruction based on memorizing without understanding has become obsolete. A better and deeper understanding of science enables students to explain phenomena and solve real-life problems, while engaging more female students (Wan, 2021). Collaborative projects and environments are particularly helpful for female students (Wang, 2012), as it can provide them with real-world applications of science, reinforcing their decision to persist (Margolis et al., 2000). When faculties embrace these real-life situation teaching techniques, female students' learning and confidence levels improve (Hyde and Gess-Newsome, 2000), nevertheless STEM majors have not yet fully embraced these more collaborative teaching styles (Laird et al., 2007). However, it is noteworthy to remark that, contrary to expected, recent studies (i.e., Sax et al., 2018) have found that feeling supported by the computing department, as well as by peers, results to be central to fostering women's and minority students' sense of belonging in the field of computing, even more than specific inclusive pedagogical practices.

The present study

While most of the existing studies addressing these topics have been conducted using mostly quantitative methods (Arriaga et al., 2011; Bernardo et al., 2017), there is a lack of qualitative research providing in-depth analysis of the views of female drop-out students. When so, although they present interesting research contributions, they have some shortcomings. This is the case of the one of Casanova et al. (2021), which is not focused on the attrition gender perspective. Others such as Chou and Chen (2015) give an extensive vision of current engineering female students' perceptions, but without pointing out the divergences between the ones who persist and the ones who switch or drop out. Even Madara and Cherotich (2016), who studied the challenges that face female engineering students, highlighted just the perspective provided by current students. The present study aims to give voice to female dropout students to seek the underlying reasons that lead them to leave their majors. Dropout and non-dropout female engineering students face similar challenges in male-dominated majors, however, there are triggers that make them decide to switch, while others decide to stay. Understanding how women who continue in engineering do differ from those who leave could help us to find useful individual and organizational tools for helping them to stay. On the other hand, we aim to get a better understanding of how female students who drop out feel that their sense of belonging and motivation to persist was undermined in a male-dominated major. Therefore, we propose the following research questions to be answered by in-depth interviews addressing the following research questions:

- RQ1: In what ways do women who drop out have different goals that lead them to consider engineering majors to be less aligned with female (communal) gender roles than women who do not drop out or men who drop out?
- RQ2: What are the main differences found in terms of self-efficacy perception between female and male students who drop out?
- RQ3: In what ways does the sense of belongingness (chilly climate) of students who drop out differ from that of students who do not drop out?
- RQ4: In what way may the perception of the curriculum (collaborative methods and grading system) affect female motivation for persisting in Engineering?
- RQ5: In what ways does role congruity perception impact female students' attrition rates?
- RQ6: What types of interventions or measures could be taken in order to better prevent female students from abandoning their studies?

Methods

Procedure

We used a mixed-methods approach combining qualitative and quantitative analyses. The overall purpose of mixed methods studies is that the use of quantitative and qualitative approaches in combination provides a better understanding of the research problem than either approach alone (Caracelli and Greene, 1993; Creswell and Plano Clark, 2017). We applied qualitative analysis to explore the ideas, behaviors, and feelings of student participants and quantitative methods to determine the direction or extent of these insights. Greene (2011) points out several advantages of mixed methods research: complementarity (the results from one method clarify the findings from the other method), development (the results from one method help to develop the use of the other method, for example, to inform future research) and expansion (in our case, seeking to extend theories about the causes that prevent women from dropping out from STEM degrees).

According to Molina-Azorín (2016), data collection refers to the sequence the researcher uses to collect both quantitative and qualitative data. In this research, we gathered the information at the same time (concurrent design) which means that researchers seek congruent findings. Thus, in the in-depth interviews, we asked participants to quantify some nodes.

Sample

We performed a non-probability sampling method including quota sampling and snowball sampling. In a quota sampling,

10.3389/fpsyg.2022.918439

researchers develop control categories, or quotas, of population elements whereas, in a snowball sampling, participants are asked to assist researchers in identifying other potential subjects (Malhotra, 2008).

In this study, we conducted 34 in-depth interviews with students participating (or that have participated) in an Engineering major and we posed 3 quotas: 10 dropout male students; 10 dropout female students, and 10 non dropout female students, however, we decided to extend the number of female non-dropout participants to strengthen the recommendations to persist (RQ6). Snowball sampling design was applied to identify potential subjects in each quota. We initially contacted 5 women who were studying for an Engineering degree and asked them to look for other women persisting in Engineering or women or men who have abandoned their Engineering studies. Looking for students who were willing to participate in the research has not been an easy matter, especially for women who drop out of engineering majors because, in most cases, they have assumed it is a personal failure that they find difficult to talk about. Therefore, we reached an agreement with the Royal Academy of Engineering and ASTI Talent and Technology Foundation, which helped us by providing the contact details of women studying engineering. Finally, the sample consisted of 9 male dropout students, 10 female dropout students, and 15 female non-dropout students. Participants were born between 1994 and 2003. All of them started an Engineering major and most of the students who dropped out changed to majors in the social sciences such as business administration, economics, business intelligence, or international relationships. Participants' characteristics are reported in Supplementary Table 1.

Study design

The research team designed the study by generating hypotheses about possible causes and associated features that prevent women from dropping out of STEM majors, based on the theoretical model developed in the previous section. This led to the design of a semi-structured interview (Marshall and Rossman, 2014) to understand the factors that influence female students' retention in engineering. Questions related to the persistence of students in engineering degrees were developed based on previous research on female students in STEM (González-Pérez et al., 2020).

To ensure the objectivity of the interview process, the authors carefully wrote and rewrote all the questions (consulting with outside third-party colleagues) both to improve construct validity and to ensure that the authors did not lead respondents in their answers (Gibbert and Ruigrok, 2010). A common set of questions was presented to all participants in a semi-structured interview to identify both positive and negative experiences that have occurred over the course of the respondents' academic life. Interviewers established a climate of trust to ensure that respondents felt safe in sharing their experiences. Thus, their experiences, rather than the authors' perspectives, drove the research.

Participants who agreed to participate in our study were scheduled for an interview with a researcher. Given the sensitive nature of the topics covered in our interviews, interviews began with an explanation of the purpose of the research, a reiteration of the assurance of confidentiality, and an opportunity to allow respondents to ask any questions before starting. All interviews were conducted online by Teams or similar apps. Online interviews, in addition to saving time in commuting, have been shown to produce as reliable information as face-to-face interviews and, in some cases, may even ease respondents' anxiety (Salmons, 2015).

Each interview lasted half an hour on average, was recorded by the interviewer, transcribed by the research team, and completely anonymized. Data were collected from December 2021 to February 2022.

Measure

Transcriptions of all the interviews were entered into Nvivo 12 to organize and manage the data. Interview questions focused on the following areas: motivations for choosing an Engineering degree, course design and subjects they feel more comfortable with, self-efficacy perception, belongingness, chilly climate with classmates and/or teachers, and socio-economic status. Indirectly, we looked for the factors underlying the women's decision to leave engineering degrees/majors, that contribute to the persistence of gender inequalities in STEM fields. The interview guide, based on the research questions and a review of the literature, included the following general questions: (1) What motivations led you to choose an engineering degree? (2) To what extent have you used collaborative projects or with a practical approach? (3) Did you feel at any time that you were not capable of getting an engineering degree? (4) How do you think relationships with peers and teachers influence the decision to persist? (5) Do you consider that socioeconomic status influences to study engineering? Nodes include gender congruity, goals (attitude and early experiences, role models, socio-economics status), self-efficacy, sense of belongingness (chilly climate with classmates and chilly climate with teachers), curriculum perception, and persistence. Nodes are described in Supplementary Table 2.

To homogenize coding methods, four interviews were randomly selected and independently analyzed by three authors to identify the content representative in each node, as well as novel themes. After coding, these three authors discussed the nodes and paragraphs representing them and agreed on node labels and definitions, developing a codebook that facilitated reliability among raters. The remaining 30 transcripts were then coded separately by two authors using the codebook and labelling segments of text according to whether the content appeared to pertain to one or more of the defined nodes.

The two coding authors then compared their individual assessments. The reliability of the coding between the authors resulted in 97.04% agreement. To test this interrater reliability, we obtained the Cohen's Kappa coefficient, resulting in 0.55. The Kappa coefficient is a quantitative measure of reliability for coders rating the same phenomena, corrected for how often the raters may agree by chance (Cohen, 1960). As Cohen suggests, a Kappa coefficient superior to 0.41 should be acceptable.

To meet the assurances on confidentiality given to participants, the authors did not involve a third party in coding interviews.

According to the quantitative research, we used the Mann–Whitney U test for independent samples (a non-parametric alternative to paired *t*-test) to rate differences in the mean values between male and female students who have dropped out; and between female students who persist and female students who have dropped out.

For this purpose, in the interview, we asked participants questions such as: (1) Do you think that men and women have different motivations to choose a university degree? (2) Did you feel more interested in subjects with a more practical content or a more collaborative approach? (3) Have you felt that you had a low self-efficacy perception? (4) Did you experience a chilly climate with teachers and/or classmates? (5) To what extent do you consider that socioeconomic status is a barrier to studying engineering?

Participants answered to the questions using an 11-point Likert scale, where 0 means that they totally disagree with the question posed and 10 that they totally agree with it. For example, in the question "have you felt that you had a low self-efficacy perception?" a rating of 10 means that he/she agrees with this negative self-view, whereas a rating of 0 refers to the opposite, showing a positive self-concept to finish his/her STEM degrees. Previous scholars have used an 11-point Likert scale to measure these issues (Nicolaidou and Philippou, 2003; Mohd Dzin and Lay, 2021; Hitches et al., 2022) whereas other researchers state that the reliability of scales increases with the number of points used (Scherpenzeel and Saris, 1995; Scherpenzeel, 2002).

Statistical analysis of the data was performed with IBM SPSS (version 27) statistical software for Windows; with a margin in the level of accuracy of 95% and an error level of 5% (statistical significance level of $\alpha = 0.05$).

Results

This research begins with an exploratory qualitative approach, followed by a quantitative analysis of the preliminary results obtained in the interviews. The use of a mixed-method research plays an important role because results obtained from both qualitative and quantitative methods enrich our understanding of the problems and questions of our research topic (Creswell, 2009; Molina-Azorín, 2016).

Qualitative analyses

Our study reveals that female students who drop out do not find practice-oriented subjects or collaborative projects in the first course, feel a lower self-efficacy perception than male students, and agree with the idea that Engineering majors fit better with male gender roles much more than non-drop out female students and, sometimes, experience a chilly climate in the classroom.

After each quote, in brackets, we have noted the number of the interview and if it is a male or a female who drops out (OUT) or persists (IN). For example (I19_Female IN), corresponds to interview number 19 which is a female who persists in an Engineering major. The encoding density, for females who drop out, females who persist, and males who drop out, can be appreciated in Supplementary Figures 1–3.

Role congruency perception

Female students who remain in these majors think that horizontal gender segregation is something that we, as a society, have overcome. Furthermore, women who did not drop out have found that they can make an impact in society through Engineering.

> "It does not have to be like this. Each person has their own goals and motivations and can do anything that he or she wants, right now, in the middle of 2022, in the middle of the 21st century. I believe that everyone can choose what they want to do and visualize themselves in one way or another and choose their path from there, without considering if they are men or women." (I19_Female IN)
>
> "I believe that, in any profession, there are aspects that can be achieved to make an impact in society… as an engineer, I believe that I can make a real impact in our society and contribute with very good achievements without being a doctor or nurse." (I08_Female IN)

However, female students who withdraw tend to think that these fields are less aligned with female gender roles, even having chosen them in the first place.

> "I think that as women, we are more focused on taking care and worrying about others […] There is a social norm that assumes women must take care of people: disabled, children, or even little brothers or sisters. Family care always falls on us and men can dedicate themselves to reaching success. So yes, I think it has something to do with having different aspirations." (I29_Female OUT)
>
> "It is true that the vast majority of men are more focused on being successful, and I think that my family and friends assumed that since I did volunteer work, I was going to choose a major related to care, such as social work, and by the time I wanted to change, I found that everyone expected me to quit Engineering." (I32_Female OUT)

Male students are equally aware of the different roles, aspirations, and goals assigned to men and women. They find that society assumes that the most demanding and competitive careers, such as Engineering, are not attractive to women and will remain in male fields.

> "That depends on how much each person has been influenced by the gender roles that have been established. Society will tell you if you are a woman, to become a nurse, for example. And there is ninety-odd percent of nurses, but my brother is a nurse. In other words, it is not to a certain extent what a person really wants, because people are also conditioned by society norms." (I23_Male OUT)
>
> "Absolutely. There are many studies that corroborate it, women have personality traits like compassion, empathy, and more focus on people. However, men generally tend to focus more on objects and tend to be more competitive and technical. And it will always remain this way. Although they may try to set quotas, men will tend to choose more technical majors, especially Engineering, as women will tend to choose majors whether related to literature, nursing or focused on caring. (…) Women and men have different motivations, of course, from a psychological, biological, and sociological point of view." (I34_Male OUT)

Having high educated parents (Social Economic Status) positively influenced choices not congruent with gender roles for some students, as mentioned by the following interviewee:

> "And with the help of my parents, who are university professors, they helped me clear my head, putting together my concerns, and the things I liked, suggesting that Engineering could be the best choice for me." (I06_Female IN)
>
> "Both my parents have Law majors, and they told me that it was a very good career choice, but as soon as I told them that I was interested in Physical Engineering they were also delighted and supported me." (I01_Female IN)

Goals

Four female students who persist and four men who have dropped out refer to early experiences in science as a relevant factor for choosing Engineering. None of the female students that withdrew referred to this type of previous experience. On the other hand, six female students who persist in Engineering refer to having been influenced by role models, compared with just one woman and one man who dropped out. Therefore, learning vicariously from role models appears to be a powerful tool resulting in higher motivation for female students to pursue and persist in Engineering.

We find out that women who persist in STEM majors tend to have had early experiences and role models that have helped them to build strong goals and motivation to pursue these studies. These influences seem to have reduced their gender stereotypes and made them feel more aligned with these male-dominated majors and roles.

> "Well, since I was a little girl, I have always liked science, I asked for gifts of chemistry sets for Christmas. I have always been good at maths and physics at school … I have been given books on why things happen, and why natural phenomena happen, which also helps a lot to be interested in science. And my family. In my family, there are several engineers, my uncle, my grandfather, my cousins, who have been real role models to me…" (I05_Female IN)
>
> "People always tell me that I should have studied medicine because I am a very curious person and I am always asking what everything means, but I think that my curiosity fits much more with getting an engineering degree because in the end you learn from so many things and explain many realities that we have around us." (I09_Female IN)

Therefore, having contact with science or role models in the early years appears to help women to reduce their gender stereotypes, allowing them to broaden their horizons and consider other types of majors not necessarily aligned with gender roles.

On the other hand, we find that female students who dropped out of engineering majors usually have not had these early science experiences or the influence of a significant role model. We have found that these women tend to choose these majors based more on agentic values, such as having a better job in the future or earning more money. These goals and motivation could not be sufficient to persist when a difficulty appears, as they are not congruent with what is expected from a woman.

> "I thought that Engineering was a career that had many professional opportunities because you can do Engineering and work as an engineer or join a consulting firm and dedicate yourself to the business world because in the end it gives you some maths tools and makes it much easier for you to get into any work (…) And that was what I liked about Engineering, I loved the idea of having an advantage over other candidates who could apply for a job […] I have never seen myself working in the Engineering world, I have always had the business world in my head." (I28_Female OUT)
>
> "In my case, I correspond to the profile of a person who wants to achieve success in life, not for getting recognition, but on a personal level. In other words, I am a very demanding person, and I knew that I wanted to pursue a career that would be interesting and lead me to have a job that I liked and where I could advance and grow." (I26_Female OUT)

On the other hand, also men students who have withdrawn from Engineering seem to be guided by agentic values, congruent with their gender roles.

10.3389/fpsyg.2022.918439

> "I was guided by the idea that by studying Engineering you will have more open doors, or you will have a greater variety of possibilities, or even being able to choose more types of paths if you want to change at a given moment." (I13_Male OUT)
>
> "It was a challenge for me, while other careers did not challenge me at that moment […] I decided to do Engineering because it was starting with the most difficult major, even not having a very clear idea what I wanted, and leaving sometime later to start other paths." (I25_Male OUT)

It seems that women who stay in Engineering have found intrinsic motivation. However, both women and men who have dropped out mimic male traits based on extrinsic motivation, which does not seem enough for helping them to persist in the major.

Self-efficacy perception

Low self-efficacy is one of the strongest barriers that women face. This low self-efficacy keeps women on the back foot in engineering majors, limits their aspirations, and leads them to feel they do not suit them. This feeling of low self-efficacy stands out especially in female students, regardless of whether they have dropped out or not. Failing exams repeatedly, having many tasks and exams to do, and comparing themselves with others are some of the arguments put forward by the participants:

> "(…) except for the first exam, what I do is fail and even though you have to dedicate many hours to it, and you have to study a lot, I felt that I did not get ahead (…) Yes, I felt frustrated and compared to my classmates, maybe, I don't know, I didn't understand very well why they were getting good marks and I couldn't." (I17_Female OUT)
>
> "There comes a time when you have so much pressure, so many things that something in your head tells you 'I can't." (I03_Female IN)
>
> "Sometimes you were very well prepared, and you couldn't get it because it was a very high level, so, for me sometimes there was a feeling of impotence." (I08_Female IN)

Sometimes, even though they get good marks, they play this down:

> "I have a friend (a female student), for example, who got a very good grade in a subject that she was retaking, and it was like oh, well, since I retook it, well obviously I'm going to be among the best, however it was more than that: she was very good. However, she was always saying that it was because she had retaken it and, well, it was normal." (I32_Female OUT)

As one of the participants shows, low(er) self-efficacy perception has a strong relation with impostor syndrome, the psychological pattern in which one doubts one's accomplishments, which makes these women have a persistent internalized fear of being exposed as a 'fraud' (Langford and Clance, 1993). This is a self-limiting feeling (de Vries, 1990), very much in line with the role incongruity these women feel (Hernandez Bark et al., 2016) being in a world that belongs to men:

> "Impostor syndrome in a woman's life is inevitable. And more so in a world of men. It's just that it's impossible not to feel inferior when you're also getting into a mess all your life that isn't your place as such." (I29_Female OUT)

Maybe all of this is due to the fact that women self-impose higher quality standards. Not only do they have to contend with the pressure of their studies but with their own feelings. Referring to this situation, one participant mentioned the following:

> "(…) moreover, we are also generally very, very demanding with ourselves. We always try to give our best." (I09_Female IN)

Another female student missed more motivation from the university:

> "I missed having an encouraging push, that they gave me a vote of confidence. (…) someone who told me: come on XXX, you're going to do very well (…) I would have liked a greater motivation towards myself to achieve it, I think that although it would have been difficult because it is difficult and I do not deny that, however, if they had given me a greater vote of confidence, maybe I would have got it." (I26_Female OUT)

However, this feeling is not shared by their male counterparts, even though they have dropped out:

> "No, at no time. The truth is that even now (after dropping out) I see myself perfectly qualified to study and graduate in Telecommunications Engineering." (I22_Male OUT)
>
> "No, I always have the impression that if I managed to focus and get serious about it, I would have got it perfectly." (I23_Male OUT)

Curriculum perception

Harsh competitive grading systems, densely packed curricula, and a lack of teaching for conceptual understanding (Seymour, 1995; Zohar and Sela, 2003) negatively affect women undergraduates in STEM majors. Whereas, hands-on tasks, employing active learning techniques, communal, collaborative learning environments, and teaching an understanding of the social relevance of physics in their everyday worlds have a positive impact on self-efficacy (Jansen et al., 2015).

Interviewees highlight the difficulty of the subjects, the assessment of learning, the overload of work, and the lack of time to do it. Three participants express it in the following way:

> "Some subjects are almost impossible, either because they are difficult, or because there is a lot of content within the syllabus, many things to study (…) The assessment is also a hindrance, they set some minimums to pass, and it is very complicated... (…) there is a lot of theory and a lot of volume for a short period of time." (I03_Female IN)
>
> "The Bologna plan places great emphasis on doing many things throughout the year (…). You also must slow down at some point. I always had exams or homework, and I think that sometimes that was quite problematic, because you are always overwhelmed, you always have things to do." (I32_Female OUT)

Indeed, this situation led the participants to a lack of motivation because they were unable to adapt. There are numerous examples of drop out men and women who described their experience in the following ways:

> "The pressure with this new way of studying, was difficult for me." (I20_Female OUT)
>
> "One of the reasons why I got frustrated with the degree was because (studying telecommunications engineering), during the first year, we did all the programming exams on paper (he refers to not using computers)." (I32_Male OUT)

Another male student goes on saying:

> "You lose motivation because it is not oriented to the real world. (…) what I found most was how abstract and outdated I saw the ways of teaching." (I25_Male OUT)

Since in the first courses there are hardly any collaborative projects or practice-oriented subjects, women who persist especially value this type of learning:

> "Practical activities help you to get an idea of what work is like after leaving university. And obviously that helps a lot to motivate you to do your best." (I07_Female IN)
>
> "Collaborative work has been useful to deepen and put into practice what we have studied (…) it gives more meaning to work…I'm doing it for a reason… I am fighting for something." (I09_Female IN)

While another female participant who dropped out, posits:

> "In the last courses I felt more comfortable (…) since we had a lot of practice-oriented subjects, a lot of laboratories… thanks to those collaborative environments in small groups I had a closer relationship with lectures." (I32_Female OUT)

Considering the methods of assessment, as a key part of the curriculum, we found that there were mentions about low grades by seven female students who stayed in Engineering, compared with six who dropped out. Their male counterparts who dropped out from the major seemed to be less affected by this low grading as none of them referred to it as a barrier. It could be explained by women's double standards, as they judge themselves more rigorously about grades. Women entering Engineering tend to be overachievers, who have had the best grades in high school. This can cause a sense of failure and of being out of place, affecting belongingness and even self-efficacy perception. Nevertheless, it does not seem to be something specific to female students who drop out, it appears to be a barrier that needs to be overcome by female students in general.

> "Sometimes you were very well prepared, however you could not pass because they asked for an unattainable level, so, I sometimes had a feeling of impotence." (I08_Female IN)
>
> "I had a very bad time, June of last year was one of the worst moments I've ever experienced. Because I have never tried so hard to be able to get something without any success. […] Your exam is shit? I have been told that many times, many times. And my tears were falling, because you cannot tell me that my exam is shit because you do not know the work behind the exam. You do not know the work behind your zero …" (I26_Female OUT)

However, an assessment method that encourages competition over collaboration is a system where female students do not feel comfortable, as stated by interviewees:

> "And you were surrounded by guys that … were really competitive, and I felt … a bad vibe, because that doesn't work for me at all." (I09_Female IN)
>
> "Men tend to be more aggressive, and competitive (…) And I think that it is because they are more used to competing, not just academically, but in life." (I20_Female OUT)

Finally, when students were asked about their study and personal life balance, this barrier was mentioned mostly by drop out students (male and female), it looks like female students who have persisted have been able to find some kind of balance:

> "Having a life while you're in college seems very complicated to me… the amount of time you need to study was disproportionate, I felt … overwhelmed, without time for my

> own life, family and friends. It seems that it is like a part of your life was missing. In other words, it's like focusing all the time on studying, because if you don't, you won't get there… maybe it had to be more years… instead of you spending your entire life studying." (I14_Female OUT)
>
> "I am very well organized. I have to relax and have free time, maybe I'll start studying soon and then I make plans, because if I am studying all day it won't work for me. […] And yes, I think that you need to organize yourself well and to study your daily hours, your 5 daily hours." (I31_Female IN)

Sense of belongingness

Women can experience a chilly climate with teachers or with their classmates. Chilly climate stands out as a barrier which can block the route to their degrees, including feelings of isolation and intimidation, sexual harassment as well as a loss in self-confidence as they progressed through their major program (Blickenstaff, 2005). Female students generally receive less attention from teachers than their male counterparts regardless of the subject or age of students (Wilkinson and Marrett, 1985; Sadker and Sadker, 1994). Furthermore, student–teacher interactions are qualitatively different for male and female students as well, while women ask more questions than men, teachers give them less feedback (Spear, 1984; Eccles and Blumenfeld, 1985; Sadker and Sadker, 1994). Four participants express this situation in the following way:

> "There was a professor who …, it was my second enrolment in Calculus, I went to review my exam …. and professors were quite old in general … so, he saw me, a woman having failed calculus for the second time and told me that I wasn't suitable for Engineering. And in the end it takes you down. And I even considered leaving Engineering and studying something else that has nothing to do with Engineering, because they have told me precisely that I was not suitable." (I02_Female IN)
>
> "I had to do a project with a male classmate and when I asked the teacher questions, the teacher always addressed my classmate, never me, he didn't explain things to me." (I14_Female OUT)

Some female participants experienced difficulties integrating with their classmates also:

> "I have had experiences with male classmates of not speaking to me or to any of the other women in class until they realized that I had the best grades and then, suddenly I was a person with whom they wanted to talk a lot." (I32_Female OUT)
>
> "A friend of mine, XXX (female) was also very, very smart and she was very good at Engineering. So, our male classmates did not ignore her because she was intelligent and she helped a lot, but nevertheless, if they saw that you couldn't contribute, they leave you aside." (I20_Female OUT)

Even though female students who do not find any problems still feel outsiders in men's networks. They do not feel they match the masculine interests, and struggle to get into a group. According to two respondents:

> "Not only is it difficult to integrate into conversations outside class, but also in class, when groups are formed because of the things they have in common or simply because they are men … for a group project, one chooses friends (…) that closeness is difficult to have in the group or even the confidence to comment on things more calmly… for example, right now I have a project with four colleagues and several times they met to do part of the work among themselves. And they didn't tell me anything … (…) Many times you don't even feel like attending classes because you know what you're going to find, the conversations they're going to have, even small jokes…." (I07_Female IN)
>
> "The way in which men relate to each other or the interests they may have outside of university are very different from what we (women) may have or the problems we may have. (…) For me it was not the same as being with my friends (women). Many times (when we met out of university) I went with a friend (female) because I did not feel comfortable." (I06_Female IN)

And all of this affects the sense of belongingness. One participant shared the following statement referring to the competitiveness among men:

> "I didn't like that atmosphere of competition and comparison of grades." (I20_Female OUT)

Others refer to the organizational system, as the university insists on the importance of changing your mind and thinking like an engineer, however some felt that they did not fit into that claim and that the system should change:

> "I felt very alone, I felt that we had to know everything (…) I didn't find any kind of support, no matter how much I asked for it." (I21_MaleOUT)
>
> "(…) I think that the system should change for those people who don't fit in." (I26_Female OUT)

In any case, both drop-out and non-drop out female students agree on the importance of having a good group of classmates at the university that supports you to continue your studies:

> "At university I have a very good group of friends with whom I do my homework, study and attend classes. We help and support each other when we don't understand something, an exam goes wrong or if we are missing some notes… It helps a lot to share worries, successes and failures. This makes studying and everyday life more enjoyable and easier." (I05_Female IN)

> "I met women and some guys with common interests and that helped me a lot. I think it is important to be comfortable at university. (…) in general, you must really want to persist and as much as you like a major, if you are alone, you feel alone, and it is very difficult for you to take it forward." (I32_Female OUT)

Persistence

Analyzing the reasons that female and male students posit for dropping out, we found that there were important differences. Women tend to argue that they have faced mostly psychological pressure and emotional barriers.

> "For me they were psychological barriers, because I left school with good grades and even though I was warned that Engineering was very hard, well I entered and basically except for the first exam, I failed everything. Even dedicating many hours and studying a lot, I felt that I was not able to go forward and that you need a huge capacity for sacrifice and being very smart to be able to graduate at the end." (I17_Female OUT)

> "You must expect the failure. This sounds very hard, but it's even harder to get through your first exam having studied 8 hours a day and get a 2 or a 1. And sometimes we are not psychologically prepared for it, especially because in high school when we study, we pass and, in the university, it is not like that. What they do not tell you at the beginning of the major is that even having studied, you are going to fail." (I16_Female OUT)

However, male students tend to cite their reasons for dropping out as not being sufficiently motivated, they never mentioned feeling they were not capable. In fact, men posit that they could cope with this high level of difficulty if they had made the effort. The problem for them was that the content was not what they have expected.

> "The problem was the work that was going to be performed after graduation because Engineering is extremely difficult, but it's nothing you can't do, if you're serious about it. But it's just that so much work, so much effort to make a piece of metal." (I23_Male OUT)

> "I think it's because of motivation, what I'm telling you is that it was difficult to find a way to start studying things that are not very attractive for most people. And realize that you don't want to do that anymore. More than you can't, it's just that you don't want to." (I25_Male OUT)

Interventions or measures to pursue

Female interviewees who have persisted in Engineering shared different measures or recommendations that have helped them. Six of them refer to emotional support from family and friends:

> "The support of your family, friends, people you trust, who encourage you to keep trying and not giving up, I think that in my case it has been the most important thing." (I06_Female IN)

Seven students also mentioned the importance of having a support network within the university: friends and colleagues with whom you can share your problems, your failures and your successes. Students posit that this allows you to not feel alone and be constant and persistent, maintaining the pace of such a demanding major.

> "At the university I have a very good group of friends with whom I do homework, study and go to class. We help and support each other when we don't understand the subjects, an exam goes wrong or if we have missed some notes. It helps a lot to be able to share worries, successes and failures. This makes studying and everyday life more enjoyable and easier." (I05_Female IN)

> "The most important thing is to count on your classmates because we have done teamwork, asked questions and I think it is super effective because they are people who are available practically 24 hours a day, you can write them whenever you want and there is always a classmate who is super smart and will know how to solve any problem or you will be able to help others with something that they have not understood, that is, you will always find someone who will be able to help you."(I27_Female IN)

Several personal qualities were also mentioned that can be helpful to succeed in Engineering. Among them, self-confidence was brought up by five students, while optimism and hope were other valuable qualities that were mentioned.

> "It is very important, of course, not to lose hope, because this is a long-distance race and to hold on and finish it you need hope and to think that you can do it." (I07_Female IN)

> "I would tell them that they can do it and that they should not believe that the male student next to them in class is smarter or that he can do better than them." (I09_Female IN)

10.3389/fpsyg.2022.918439

Quantitative results

In the interviews, we asked participants to quantify some of the nodes. Mann–Whitney U test was used to analyze differences in the mean values between male and female students who have dropped out (see Table 1) and between female students who persist and females who have dropped out (see Table 2).

While comparing quantitative results from male and female students that have dropped out, no significant differences were found. Thus, even though it might seem that the main reasons for dropping out might have been similar, the underlying insights seem to have affected them differently. Whereas the comparison between female students who persist and females who have dropped out shows significant differences in goals, curriculum perception, and sense of belongingness.

Regarding goals, the results highlight that there are differences in the perception of gender role congruity perception between female students who drop out and those who do not. Female students withdrawing from Engineering find that there are greater differences between interests and aspirations (goals). Even having chosen this type of major at first, these women end up thinking that there are still different career paths for women and men. The question that arises is if they have always considered that there are careers more suitable for women or men or if the experience in a male-dominated major has led them to this perception.

Concerning curriculum perception, significant differences were found since non-drop-out female students seem to prefer practical and collaborative subjects. However, this can be explained by the fact that the interviewees state that this practical content is taught in the last years of the major. Therefore, students who have dropped out may not have had the opportunity to learn this practical content.

In the case of self-efficacy perception, no significant differences were found between self-efficacy perception for female students who persist and those who do not.

Lastly, in the case of sense of belongingness, marginal significance was found. Drop-out students found this climate much more difficult than their peers who have not dropped out. This is consistent with the qualitative findings as having a network of friends in class is mentioned as one of the key elements to persist.

TABLE 1 Mann–Whitney U Test on dropped out students (male vs. female).

	Gender		
	Male	Female	Mann–Whitney U Test
Goals	5.72	6.80	39.50
Curriculum perception	7.50	6.22	25.00
Self-efficacy perception	4.94	5.85	33.50
Sense of belongingness	2.44	4.22	29.50

$^{**}p < 0.01$; $^{*}p < 0.05$.

TABLE 2 Mann Whitney U Test on female students (non-drop out vs. drop out).

	Persistence in the major		
	Non-drop out	Drop out	Mann–Whitney U Test
Goals	4.63	6.80	37.00**
Curriculum perception	8.17	6.22	34.50**
Self-efficacy perception	3.90	5.85	47.00
Sense of belongingness	1.60	4.22	39.5*

$^{**}p < 0.01$; $^{*}p < 0.05$.

Conclusion and discussion

This research contributes to the literature (Wolffram et al., 2009; Sweeney, 2020; Mickelson et al., 2022) on amplifying Tinto's persistence model from a gendered perspective. The findings of the present study make several important contributions to the existing literature on persistence in engineering majors, which can help future research and policies on this topic. Much of the previous research focused on reasons to persist without considering gender bias. Understanding how gender stereotypes and roles congruity affect female persistence can help to design better and more effective interventions.

Although Tinto's model has received different critiques in the last years, with authors proposing its extension to include specific facets that might affect to concrete population (i.e., Botanga et al., 2021, with URM), our results support the theoretical model as a general framework to understand women drop out from Engineering when including the gender perspective. In this line, our results show that all the theoretical constructs proposed by the model (i.e., goals, curriculum perception, self-efficacy, sense of belongingness) work as (des)motivators in the expected way. Besides, although some results appear to be non-gendered (for instance, intrinsic motivators work better for both men and women to persist on the major), most of them are clearly gendered. In this line, most of the reasons of why women drop-out can be explained from the social role theory (Eagly, 1987a,b) as mainstreaming in each of the Tinto's model constructs. Thus, no having contact with science or Engineer women that act as role models in early years affect female students both in establishing goals and in the sense of a low self-efficacy perception; the role-incongruity perception between being women (communal goals) and studying Engineering majors (agentic goals) affect not only to women themselves (increasing the impostor syndrome) but also to peer and teachers' support, which increases the chilly environment and, in return, decrease their sense of belongingness. In this line, drop-out women still consider that there are still different career paths for women and men (different goals). Also, there are differences at the curricula, as female students who drop out do not find practice-oriented subjects or collaborative projects in the first course that could serve as mastery experiences, which leads to a decrease of their self-efficacy.

Thus, in global, all this results in the fact that women tend to argue that they have faced mostly psychological pressure and emotional barriers, whereas men never mentioned it.

This research will allow to implement effective interventions to increase women's persistence in engineering majors. The study advances our understanding of the barriers and obstacles that face female students in Engineering showing that motivation to persist is the result of multiple factors (i.e., goals, self-efficacy, curriculum, sense of belonging) that are affected by gender role perceptions. In the following sections, we will point out the most remarkable results and relate them with its practical implication in form of intervention proposals.

Role congruency perception and role models

Our findings suggest that role congruency perception and lack of role models are more pronounced in female students who drop out. They also suggest that students know nearly nothing about the practical applications of Engineering when they enroll at university. Thus, it would be important to have interventions in early years, when gender stereotypes begin to affect expectations, interests and academic choices. Role models have been shown by extensive prior literature to be critical in motivating students to follow a path and achieve goals (Collins, 1996; Lockwood and Kunda, 1997). However, there is no unanimous agreement on how these models should be. Certainly, they must be seen as competent and successful by female students (Marx et al., 2013). Likewise, it is essential that they can feel identified, so they must belong to the same gender and ethnicity (Lockwood, 2006). Regarding lack of knowledge about the content of Engineering and its practical application, we recommend holding sessions in high schools for both female and male students (Falco and Summers, 2019). These sessions could be held by current Engineering female students, because it will be easier for high schoolers to identify with them (Mussweiler, 2003). Relating to the content, we propose that these role models focus on how these professions can contribute to society, showing innovations congruent with communal goals led by female engineers (Boucher et al., 2017). It is important to highlight that these professions solve real problems, can help others and improve our lives (González-Pérez et al., 2020). We recommend holding these interventions for both female and male students as it is also important for male students to dispel gender stereotypes. Nevertheless, these sessions could also be held by men not conforming to agentic masculine stereotypes, which could help women and girls to see them as allies (Cheryan et al., 2011). Furthermore, university summer sessions for high schoolers could be a game-changing early experience to boost interest in Engineering (Kitchen et al., 2018). To reduce these gender stereotypes, it could be interesting to have informative sessions with high school teachers and families to broaden their minds, as parents and teachers become principal role models, advisors and supporters for females' academic choices (Gunderson et al., 2012). Furthermore, we suggest inviting career advisors at high school level, as they could unbiasedly help students to find what they are good at and what career paths are more suitable for them regarding their interests and strengths (Falco, 2017).

Self-efficacy perception

A second conclusion of our research is that female students, regardless of whether they have dropped out or not, tend to have lower self-efficacy perceptions than males, which can lead them to drop out. Female students often self-impose higher standards of excellence, making them believe that they are not capable of persisting. In a context where women are a minority and, as noted, face multiple barriers, extensive previous research has identified self-efficacy as a key predictor of women's success in engineering (Blaisdell, 2000; Marra et al., 2005). To promote self-efficacy perception, it is worthwhile focusing on the self-efficacy sources proposed by Bandura (1997). So, following Betz (2004), we propose intervention programs that would include: (1) activities in which they can recognize that they are/were successful (mastery experiences), (2) interaction with female role models (both professionals and recent students) through mentorship programs (social modelling), (3) professors focusing on students' success and capability to perform difficult tasks (verbal persuasion), and (4) promoting positive emotions through both interventions to foster a growth mindset highlighting that intelligence and ability are not fixed traits and positive psychological programs (emotional and physiological states) to positively influence their criteria to judge their capability and vulnerability. These interventions can help to demonstrate that hard work and effort can help female students to overcome challenges, and that they are able to do it.

Sense of belongingness

Sense of belongingness becomes a key predictive factor in terms of persistence (Walton et al., 2015). According to our results, both drop-out and non-drop out female students agree on the importance of having a group of classmates at university and a family that support you to continue your studies. Literature has identified that the sense of belonging is related to the skills to make an effort in the face of difficulties (Vaz et al., 2015). These skills become even more relevant in a context such as that of engineering students who need to overcome important barriers and, on many occasions, feel alone and without support (Strayhorn, 2018). On the one side, having a strong supportive network of peers motivates female students to not drop out (Limbert, 1995). Therefore, we propose mentorship programs within the university where freshman engineering students are mixed with sophomore, junior or senior engineering students (Packard, 1999) with the aim of retaining female students through a nurturing mentoring program, designed to build a network with other female Engineering students with whom they can easily identify (Stout et al., 2011). Thus, these programs will enhance personal support through contacts with peer female role models, will build confidence and self-efficacy as mentees will see that their mentors have been able to succeed in Engineering and provides valuable emotional support. This mentorship program can also involve collaboration between university students and networks of engineers, in order to help seniors in their immersion in the professional workplace. Another intervention could be peer-led team (Horwitz et al., 2009) learning to provide female students with an efficient and supportive study group, where through workshops, a coached student who has previously been successful in the course facilitates learning. It could also be interesting to promote interventions with male and female students to highlight the importance that diversity has for innovation, promoting mix-gendered groups. Finally, student-run clubs and initiatives can also enhance a sense of belonging (Sahin, 2013).

On the other side, as it is crucial for a sense of belongingness to feel encouraged and motivated by the faculty, the Gender Compliance Committee or Diversity and Inclusion Dean should have more significance. Communications, performance and language to detect gender stereotypes need to be carefully reviewed to mitigate gender stereotypes (Cheryan et al., 2011). Following Blickenstaff (2005), course materials and assignments should also be reviewed to add female scientists and their achievements to shift perceptions about who belongs, while promoting diversity-related activities. Furthermore, training or workshops with the faculty to foster some self-reflection, review performance, identify gender bias and implement solutions could be another interesting intervention.

Curriculum perception

Another finding of our research are the masculine biases in the curriculum that sometimes prevent female students from persisting. Its content is adapted to the interests and perspectives of both the teacher and the dominant social group in the class, or both (Beder, 1989; Lewis, 1995). This leads to a new difficulty for engineering students since, since most of the faculty and engineering students are men, they may feel uncomfortable or excluded in class. In short, the content of the curriculum becomes a new barrier that can lead students to drop out, switch or not succeed in their majors. Therefore, we propose active learning and project-based instruction using collaboration techniques from the first years (Zastavker et al., 2006; Dominguez et al., 2019), to enhance a sense of community rather than a competitive environment. Incorporating service-learning projects to promote the idea that Engineering helps to improve society and allowing more choice in terms of subjects could boost interest and motivation. These new subjects could be more focused on social purposes, environmental impact or sustainability ethics that connect better with communal goals (Diekman et al., 2015).

Apart from the curriculum itself, our research found masculine biases in the grading systems. We suggest increasing collaboration rather than competitiveness. Instead of multiple-choice tests or exams where only the final answer matters, we propose replacing them with open-ended evaluations. A constructive response system that allows students to show their competence through writing has been proved to be more suitable for female students (Weaver and Raptis, 2001).

Conclusion

Finally, as a general intervention for fostering persistence, we suggest highlighting the importance of building soft skills, as mentioned before in the case of self-efficacy. Self-efficacy, hope, resilience and optimism (positive psychological capital; Luthans et al., 2007) are qualities that can help female students to overcome the obstacles and barriers they will find in their academic and professional careers. On the other hand, having counselling and psychological services could be a useful tool to reduce stress, anxiety and depression among female students.

In conclusion, based on the findings of this research, including a gendered perspective in Engineering fields provide a promising route to retaining female students. With our empirical results, we have been able to validate our proposed theoretical framework and build upon each of the parts of Tinto's well-known validated theoretical model of persistence, incorporating this gender perspective. Women who drop out of Engineering highlight in the interviews that goals incongruity leads them to low levels of motivation, affecting persistence. The results show that this lack of congruity influences mainly belongingness and self-efficacy perception. There are several practices that institutions should revisit and rethink to provide the necessary support to Engineering female students who are struggling. According to the gender differences outlined in this research, we cannot understand women's persistence in engineering without a gender perspective. Women enter male-dominated majors where they do not feel as if they belong; for instance, they stated that they are more comfortable responding to praise than to challenge. We have found that including these communal goals for real could improve retention in these majors, as they could be seen as congruent with their priorities. Findings from the present study allow policymakers and organizations to implement interventions which encourage female student persistence in male-dominated fields. Providing women with a strong support system can help them to prevail over barriers which they may face during their Engineering education. All these measures should be accompanied by a learning and social environment that promotes the reduction of gender stereotypes (Solbes-Canales et al., 2020), so the next generation of potential female engineers believe that they will be successful.

Limitations and directions for future research

These results are based on a limited sample of female engineers who have dropped out. A larger sample would be desirable, especially to strength the quantitative analysis with more robust methodologies. It would also be interesting to delve into whether the reasons are common to all engineering disciplines in general or are limited to some specific ones.

The findings from the present study suggest other promising directions for future research, for example to carry out a longitudinal study. Understanding what kinds of barriers students face at different points in their careers can provide a more comprehensive view and help design effective measures.

Data availability statement

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

Author contributions

SG-P, MM, VP, and EC-G designed, performed, and analyzed the research, searched the literature, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research has received grant PID2020-114183RB-I00, funded by MCIN/AEI/10.13039/501100011033. This research has received financial support from: Cátedra Universidad CEU San Pablo and Mutua Madrileña, Spain (060516-USPMM-03/18) and from Karmo Spirit S.L (Ref.: 001). The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

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

Acknowledgments

We acknowledge the generosity of Real Academia de Ingenieria and ASTI Talent and Technology Foundation without which the present study could not have been completed. We are also grateful to Professor Ruth Mateos de Cabo for their valuable comments and suggestions.

Supplementary material

The supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.918439/full#supplementary-material

References

Arriaga, J., Burillo, V., Carpeño, A., and Casaravilla, A. (2011). "Caracterización de los tipos de abandono. Dividamos el problema y venceremos más fácilmente" in *I Conferencia Latinoamericana sobre el Abandono en la Enseñanza Superior*. ed. J. Arriaga (Madrid: Dpto. de Publicaciones de la EUIT de Telecomunicación), 75–82.

Astin, H. S., and Sax, L. J. (1996). "Developing scientific talent in undergraduate women," in *The Equity Equation: Fostering the Advancement of Women in the Sciences, Mathematics, and Engineering*. eds. C. S. Davidset al. (San Francisco, CA: Jossey-Bass), 96–121.

Ayre, M., Mills, J., and Gill, J. (2013). 'Yes, I do belong': the women who stay in engineering. *Eng. Stud.* 5, 216–232. doi: 10.1080/19378629.2013.855781

Bakan, D. (1966). *The Duality of Human Existence: An Essay on Psychology and Religion*. Skokie, IL: Rand McNally.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *Am. Psychol.* 37, 122–147. doi: 10.1037/0003-066X.37.2.122

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Dev. Psychol.* 25, 729–735. doi: 10.1037/0012-1649.25.5.729

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: W H Freeman & Co.

Beasley, M. A., and Fischer, M. J. (2012). Why they leave: the impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Soc. Psychol. Educ.* 15, 427–448. doi: 10.1007/s11218-012-9185-3

Beder, S. (1989). Towards a more representative engineering education. *Int. J. Appl. Eng. Educ.* 5, 173–182.

Bensimon, E. M. (2007). The underestimated significance of practitioner knowledge in the scholarship on student success. *Rev. High. Educ.* 30, 441–469. doi: 10.1353/rhe.2007.0032

Bernardo, A., Cervero, A., Esteban, M., Tuero, E., Casanova, J. R., and Almeida, L. S. (2017). Freshmen program withdrawal: types and recommendations. *Front. Psychol.* 8:1544. doi: 10.3389/fpsyg.2017.01544

Betz, N. E. (2004). Contributions of self-efficacy theory to career counseling: a personal perspective. *Career Dev. Q.* 52, 340–353. doi: 10.1002/j.2161-0045.2004.tb00950.x

Blackburn, B. (2017). The status of women in STEM in higher education: a review of the literature 2007–2017. *Sci. Technol. Libr.* 36, 235–273. doi: 10.1080/0194262X.2017.1371658

Blaisdell, S. L. (2000). *Social Cognitive Theory Predictors of EntryI Engineering Majors for High School Students*. Tempe, AZ: Arizona State University.

Blickenstaff, J. C. (2005). Women and science careers: leaky pipeline or gender filter? *Gend. Educ.* 17, 369–386. doi: 10.1080/09540250500145072

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

Blumenfeld, P., Fishman, B. J., Krajcik, J., Marx, R. W., and Soloway, E. (2000). Creating usable innovations in systemic reform: scaling up technology-embedded project-based science in urban schools. *Educ. Psychol.* 35, 149–164. doi: 10.1207/S15326985EP3503_2

Botanga, C., Blanc, S., Jones, L. R. II, Day, M., and Charles, M. (2021). Continued STEM commitment in light of 2020 events: a perspective from the Illinois Louis stokes Alliance for minority participation. *Front. Educ.* 6:681027. doi: 10.3389/feduc.2021.681027

Boucher, K. L., Fuesting, M. A., Diekman, A. B., and Murphy, M. C. (2017). Can I work with and help others in this field? How communal goals influence interest and participation in STEM fields. *Front. Psychol.* 8:901. doi: 10.3389/fpsyg.2017.00901

Braxton, J., and Hirschy, A. S. (2004). "Reconceptualizing antecedents of social integration in student departure," in *Retention and Student Success in Higher Education*. eds. M. Yorke and B. Longden (Berkshire, UK: Open University Press), 89–102.

Brophy, J. E. (2004). *Motivating Students to Learn. 2nd Edn*. New York: Lawrence Erlbaum Associates. Available at: http://books.google.com.tr/books?id=gD7_VSNmu7cC&dq=isbn:0805847723 (Accessed July 07, 2022).

Caracelli, V. J., and Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educ. Eval. Policy Anal.* 15, 195–207. doi: 10.3102/01623737015002195

Casad, B. J., Petzel, Z. W., and Ingalls, E. A. (2019). A model of threatening academic environments predicts women STEM majors' self-esteem and engagement in STEM. *Sex Roles* 80, 469–488. doi: 10.1007/s11199-018-0942-4

Casanova, J. R., Vasconcelos, R., Bernardo, A. B., and Almeida, L. S. (2021). University dropout in engineering: motives and student trajectories. *Psicothema* 33, 595–601. doi: 10.7334/psicothema2020.363

Centro de Investigación y Documentación Educativa (CIDE), & Instituto de la Mujer [Center of Research and Educational Documentation & Women's Institute] (1988). *La Presencia de las Mujeres en el sistema Educativo [The Presence of Women in the Educational System]*. Madrid: Instituto de la Mujer.

Chavatzia, T. (2017). *Cracking the Code: Girls' and Women's Education in Science, Technology, Engineering and Mathematics (STEM)*. Paris, France: Unesco, 20.

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, 1045–1060. doi: 10.1037/a0016239

Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., and Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM? *Soc. Psychol. Personal. Sci.* 2, 656–664. doi: 10.1177/1948550611405218

Chou, P. N., and Chen, W. F. (2015). Female engineering students' perceptions of college learning experiences: a qualitative case study in Taiwan. *Int. J. Eng. Educ.* 31, 2–11.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Colbeck, C. L., Cabrera, A. F., and Terenzini, P. T. (2001). Learning professional confidence: linking teaching practices, students' self-perceptions, and gender. *Rev. High. Educ.* 24, 173–191. doi: 10.1353/rhe.2000.0028

Collins, R. L. (1996). For better or worse: the impact of upward social comparison on self-evaluations. *Psychol. Bull.* 119, 51–69. doi: 10.1037/0033-2909.119.1.51

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 3rd Edn.* Thousand Oaks, CA: Sage Publications.

Creswell, J. W., and Plano Clark, V. L. (2017). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage Publications.

D'Amico, M. M., Dika, S. L., Elling, T. W., Algozzine, B., and Ginn, D. J. (2014). Early integration and other outcomes for community college transfer students. *Res. High. Educ.* 55, 370–399. doi: 10.1007/s11162-013-9316-5

de Vries, M. F. K. (1990). The impostor syndrome: developmental and societal issues. *Hum. Relat.* 43, 667–686. doi: 10.1177/001872679004300704

Delaney, J. M., and Devereux, P. J. (2019). Understanding gender differences in STEM: evidence from college applications. *Econ. Educ. Rev.* 72, 219–238. doi: 10.1016/j.econedurev.2019.06.002

Diekman, A. B., Brown, E. R., Johnston, A. M., and Clark, E. K. (2010). Seeking congruity between goals and roles: a new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychol. Sci.* 21, 1051–1057. doi: 10.1177/0956797610377342

Diekman, A. B., and Eagly, A. H. (2008). "Of men, women, and motivation: a role congruity account," in *Handbook of Motivation Science*. eds. J. Y. Shah and W. L. Gardner (New York: Guilford Press), 434–447.

Diekman, A. B., and Steinberg, M. (2013). Navigating social roles in pursuit of important goals: a communal goal congruity account of STEM pursuits. *Soc. Personal. Psychol. Compass* 7, 487–501. doi: 10.1111/spc3.12042

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. *Personal. Soc. Psychol. Rev.* 21, 142–175. doi: 10.1177/1088868316642141

Diekman, A. B., Weisgram, E. S., and Belanger, A. L. (2015). New routes to recruiting and retaining women in STEM: policy implications of a communal goal congruity perspective. *Soc. Issues Policy Rev.* 9, 52–88. doi: 10.1111/sipr.12010

Dominguez, A., Truyol, M. E., and Zavala, G. (2019). Professional development program to promote active learning in an engineering classroom. *Int. J. Engin. Educ.* 35, 424–433.

Eagly, A. H. (1987a). *Sex Differences in Social Behavior: A Social-Role Interpretation*. New York: Lawrence Erlbaum Associates, Inc.

Eagly, A. H. (1987b). Reporting sex differences. *Am. Psychol.* 42, 756–757. doi: 10.1037/0003-066X.42.7.755

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

Eagly, A. H., Wood, W., and Johannesen-Schmidt, M. C. (2004). "Social role theory of sex differences and similarities: implications for the partner preferences of women and men," in *The Psychology of Gender. 2nd Edn*. eds. A. H. Eagly, A. E. Beall and R. J. Sternberg (New York: The Guilford Press), 269–295.

Eaton, Y. M., Mitchell, M. L., and Jolley, J. M. (1991). Gender differences in the development of relationships during late adolescence. *Adolescence* 26, 565–568. PMID: 1962539

Eccles, J. S., and Blumenfeld, P. (1985). "Classroom experiences and student gender: are there differences and do they matter?" in *Gender Influences in Classroom Interaction*. eds. L. C. Wilkinson and C. B. Marrett (New York: Academic Press), 79–114.

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

Eddy, S. L., and Brownell, S. E. (2016). Beneath the numbers: a review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines. *Phys. Rev. Phys. Educ. Res.* 12:020106. doi: 10.1103/PhysRevPhysEducRes.12.020106

European Institute for Gender Equality (EIGE) (n.d.). Gender perspective. Available at: https://eige.europa.eu/thesaurus/terms/1197 (Accessed July 07, 2022).

Eurostat (2021). Women in science and engineering. Available at: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20210210-1 (Accessed July 07, 2022).

Falco, L. D. (2017). The school counselor and STEM career development. *J. Career Dev.* 44, 359–374. doi: 10.1177/0894845316656445

Falco, L. D., and Summers, J. J. (2019). Improving career decision self-efficacy and STEM self-efficacy in high school girls: evaluation of an intervention. *J. Career Dev.* 46, 62–76. doi: 10.1177/0894845317721651

Fervers, L., Jacob, M., Piepenburg, J., and Beckmann, J. (2020). *No Risk, No Gain. Experimental Evidence on the Role of Risk and Return Preferences for Gender Differences in Choice of Major—And What To Do About It*. Mimeo: University of Cologne.

Flam, F. (1991). Still a "chilly climate" for women? Women in astronomy and physics say they face not so much overt discrimination as a pattern of "micro-inequalities". The remedy: more women. *Science* 21, 709–714. doi: 10.1016/0360-3016(91)90690-6

Geist, E. A., and King, M. (2008). Different, not better: gender differences in mathematics learning and achievement. *J. Instr. Psychol.* 35, 43–52.

Gibbert, M., and Ruigrok, W. (2010). The "what" and "how" of case study rigor: three strategies based on published work. *Organ. Res. Methods* 13, 710–737. doi: 10.1177/1094428109351319

González-Pérez, S., Mateos de Cabo, R., and Sáinz, M. (2020). Females in STEM: is it a female role-model thing? *Front. Psychol.* 11:2204. doi: 10.3389/fpsyg.2020.02204

Goodenow, C. (1993). Classroom belonging among early adolescent students: relationships to motivation and achievement. *J. Early Adolesc.* 13, 21–43. doi: 10.1177/0272431693013001002

Greene, W. H. (2011). *Econometric Analysis. 5th Edn.* Essex, UK: Pearson.

Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: is it the school that matters? *Econ. Educ. Rev.* 29, 911–922. doi: 10.1016/j.econedurev.2010.06.010

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

Guzdial, M., Rick, J., and Kehoe, C. (2001). Beyond adoption to invention: teacher-created collaborative activities in higher education. *J. Learn. Sci.* 10, 265–279. doi: 10.1145/572133.572137

Hall, R. M., and Sandler, B. R. (1982). *The Classroom Climate: A Chilly One for Women? Project on the Status and Education of Women*. Washington, DC: Association of American Colleges.

Halpern, D. F. (2004). A cognitive-process taxonomy for sex differences in cognitive abilities. *Curr. Dir. Psychol. Sci.* 13, 135–139. doi: 10.1111/j.0963-7214.2004.00292.x

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 Interest* 8, 1–51. doi: 10.1111/j.1529-1006.2007.00032.x

Hamm, J. M., Perry, R. P., Chipperfield, J. G., Hladkyj, S., Parker, P. C., and Weiner, B. (2020). Reframing achievement setbacks: a motivation intervention to improve 8-year graduation rates for students in science, technology, engineering, and mathematics (STEM) fields. *Psychol. Sci.* 31, 623–633. doi: 10.1177/0956797620904451

Hernandez Bark, A. S., Escartín, J., Schuh, S. C., and van Dick, R. (2016). Who leads more and why? A mediation model from gender to leadership role occupancy. *J. Bus. Ethics* 139, 473–483. doi: 10.1007/s10551-015-2642-0

Hitches, E., Woodcock, S., and Ehrich, J. (2022). Building self-efficacy without letting stress knock it down: stress and academic self-efficacy of university students. *Int. J. Educ. Res.* 3:100124. doi: 10.1016/j.ijedro.2022.100124

Horwitz, S., Rodger, S. H., Biggers, M., Binkley, D., Frantz, C. K., Gundermann, D., et al. (2009). Using peer-led team learning to increase participation and success of under-represented groups in introductory computer science. *ACM SIGCSE Bulletin* 41, 163–167. doi: 10.1145/1539024.1508925

Hyde, M. S., and Gess-Newsome, J. (2000). Adjusting educational practice to increase female persistence in the sciences. *J. Coll. Stud. Retent.: Res. Theory Pract.* 1, 335–355. doi: 10.2190/8WV7-UWY2-A1G9-7U3Y

Isphording, I. E., and Qendrai, P. (2019). Gender Differences in Student Dropout in STEM. IZA Research Report No. 87. Available at: https://legacy.iza.org/en/webcontent/publications/reports/report_pdfs/iza_report_87.pdf (Accessed July 07, 2022).

Jansen, M., Scherer, R., and Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: differential relations to antecedents and educational outcomes. *Contemp. Educ. Psychol.* 41, 13–24. doi: 10.1016/j.cedpsych.2014.11.002

Johnson, M. D., Sprowles, A. E., Goldenberg, K. R., Margell, S. T., and Castellino, L. (2020). Effect of a place-based learning community on belonging, persistence, and equity gaps for first-year STEM students. *Innov. High. Educ.* 45, 509–531. doi: 10.1007/s10755-020-09519-5

Kaldaras, L., Akaeze, H., and Krajcik, J. (2021). A methodology for determining and validating latent factor dimensionality of complex multi-factor science constructs measuring knowledge-in-use. *Educ. Assess.* 26, 241–263. doi: 10.1080/10627197.2021.1971966

Kelly, A. V. (2009). *The Curriculum: Theory and Practice.* New York: Sage Publications.

Kelly, A. M. (2016). Social cognitive perspective of gender disparities in undergraduate physics. *Phys. Rev. Phys. Educ. Res.* 12:020116. doi: 10.1103/PhysRevPhysEducRes.12.020116

Keup, J. R. (2005). The impact of curricular interventions on intended second year re-enrollment. *J. Coll. Stud. Retent.: Res. Theory Pract.* 7, 61–89. doi: 10.2190/2DCU-KABY-WVQH-2F8J

Kitchen, J. A., Sonnert, G., and Sadler, P. M. (2018). The impact of college-and university-run high school summer programs on students' end of high school STEM career aspirations. *Sci. Educ.* 102, 529–547. doi: 10.1002/sce.21332

Laird, J., Kienzl, G., DeBell, M., and Chapman, C. (2007). *Dropout Rates in the United States: 2005. Compendium Report*. NCES 2007-059. National Center for Education Statistics.

Langford, J., and Clance, P. R. (1993). The imposter phenomenon: recent research findings regarding dynamics, personality and family patterns and their implications for treatment. *Psychotherapy: theory, research, practice. Training* 30, 495–501. doi: 10.1037/0033-3204.30.3.495

Lent, R. W., Brown, S. D., and Hackett, G. (2002). Social cognitive career theory. *Career Choice Dev.* 4, 255–311. doi: 10.1177/1069072716657811

Lewis, S. (1995). *Chilly Courses for Women? Some Engineering and Science Experiences, in Women, Culture and Universities: A Chilly Climate?* Sydney, Australia: University of Technology.

Limbert, C. (1995). Chrysalis, a peer mentoring program for faculty and staff women. *NWSA J.* 7, 86–89.

Locke, E. A., and Latham, G. P. (1990). *A Theory of Goal Setting and Task Performance.* Englewood Cliffs, NJ: Prentice Hall.

Lockwood, P. (2006). "Someone like me can be successful": do college students need same-gender role models? *Psychol. Women Q.* 30, 36–46. doi: 10.1111/j.1471-6402.2006.00260.x

Lockwood, P., and Kunda, Z. (1997). Superstars and me: predicting the impact of role models on the self. *J. Pers. Soc. Psychol.* 73, 91–103. doi: 10.1037/0022-3514.73.1.91

López-Sáez, M., Puertas, S., and Sáinz, M. (2011). Why don't girls choose technological studies? Adolescents' stereotypes and attitudes towards studies related to medicine or engineering. *Span. J. Psychol.* 14, 74–87. doi: 10.5209/revSJOP.2011.v14.n1.6

Luthans, F., Avolio, B. J., Avey, J. B., and Norman, S. M. (2007). Positive psychological capital: measurement and relationship with performance and satisfaction. *Pers. Psychol.* 60, 541–572. doi: 10.1111/j.1744_6570.2007.00083.x

Ma, Y., and Liu, Y. (2015). Race and STEM degree attainment. *Sociol. Compass* 9, 609–618. doi: 10.1111/soc4.12274

Madara, D. S., and Cherotich, S. (2016). Challenges faced by female-students in engineering-education. *J. Educ. Pract.* 7, 8–22.

Makarem, Y., and Wang, J. (2020). Career experiences of women in science, technology, engineering, and mathematics fields: A systematic literature review. *Hum. Resour. Dev. Q.* 31, 91–111. doi: 10.1002/hrdq.21380

Malhotra, N. (2008). *Investigación de mercados. 5th Edn.* Mexico: Pearson.

Mannon, S. E., and Schreuders, P. D. (2007). All in the (engineering) family?—The family occupational background of men and women engineering students. *J. Women Minorities Sci. Eng.* 13, 333–351. doi: 10.1615/JWomenMinorScienEng.v13.i4.20

Margolis, J., Fisher, A., and Miller, F. (2000). The anatomy of interest: women in undergraduate computer science. *Women's Stud. Q.* 28, 104–127.

Marra, R., Schuurman, M., Moore, C., and Bogue, B. (2005). "Women Engineering Students' Self Efficacy Beliefs—The Longitudinal Picture" in *2005 Annual Conference*; June 12–15, 10–1475.

Marsh, H. W., Van Zanden, B., Parker, P. D., Guo, J., Conigrave, J., and Seaton, M. (2019). Young women face disadvantage to enrollment in university STEM coursework regardless of prior achievement and attitudes. *Am. Educ. Res. J.* 56, 1629–1680. doi: 10.3102/0002831218824111

Marshall, C., and Rossman, G. B. (2014). *Designing Qualitative Research. 3rd Edn.* New York: Sage Publications.

Marx, D. M., Monroe, A. H., Cole, C. E., and Gilbert, P. N. (2013). No doubt about it: when doubtful role models undermine men's and women's math performance under threat. *J. Soc. Psychol.* 153, 542–559. doi: 10.1080/06224545.2013.778811

Marx, D. M., and Roman, J. S. (2002). Female role models: protecting women's math test performance. *Personal. Soc. Psychol. Bull.* 28, 1183–1193. doi: 10.1177/01461672022812004

Massachusetts Institute of Technology (1999). *A Study of the Status of Women Faculty in Science at MIT: How a Committee on Women Faculty Came to Be Established by the Dean of the School of Science, What the Committee and the Dean Learned and Accomplished, and Recommendations for the Future.* Cambridge, MA: Massachusetts Institute of Technology.

Master, A., Cheryan, S., and Meltzoff, A. N. (2016). Computing whether she belongs: stereotypes undermine females' interest and sense of belonging in computer science. *J. Educ. Psychol.* 108, 424–437. doi: 10.1037/edu0000061

Mickelson, R. A., Mikkelsen, I., Dorodchi, M., Cukic, B., and Horn, T. (2022). Fostering greater persistence among underserved computer science undergraduates: a descriptive study of the I-PASS project. *J. Coll. Stud. Retent.: Res. Theory Pract.* 1–28. doi: 10.1177/15210251221086928

Mitchell, M. (1993). Situational interest: its multifaceted structure in the secondary school mathematics classroom. *J. Educ. Psychol.* 85, 424–436. doi: 10.1037/0022-0663.85.3.424

Mohd Dzin, N. H., and Lay, Y. F. (2021). Validity and reliability of adapted self-efficacy scales in Malaysian context using PLS-SEM approach. *Educ. Sci.* 11:676. doi: 10.3390/educsci11110676

Molina-Azorín, J. F. (2016). Mixed methods research: an opportunity to improve our studies and our research skills. *Eur. J. Manag. Bus. Econ.* 25, 37–38. doi: 10.1016/j.redeen.2016.05.001

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., and Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proc. Natl. Acad. Sci.* 109, 16474–16479. doi: 10.1073/pnas.1211286109

Mussweiler, T. (2003). Comparison processes in social judgment: mechanisms and consequences. *Psychol. Rev.* 110, 472–489. doi: 10.1037/0033-295X.110.3.472

Nicolaidou, M., and Philippou, G. (2003). "Attitudes towards mathematics, self-efficacy and achievement in problem solving," in *European Research in Mathematics Education III*. ed. M. A. Mariotti (Pisa, Italy: University of Pisa), 1–11.

Nicoletti, M. C. (2019). Revisiting the Tinto's theoretical dropout model. *High. Educ. Stud.* 9, 52–64. doi: 10.5539/hes.v9n3p52

OECD (2019). Stats.oecd.org. 2022. *Distribution of graduates and new entrants by field: share of tertiary graduates by field of education and gender* [Online]. Available at: https://stats.oecd.org/Index.aspx?QueryId=109881# (Accessed June 22, 2022).

Ofori-Boadu, A. N. (2018). "Improving Middle-School Girls' Knowledge, Self-Efficacy, and Interests in 'Sustainable Construction Engineering' through a STEAM ACTIVATED! Program." in *Paper presented at 2018 ASEE Annual Conference & Exposition*; June 23–27; Salt Lake City, Utah. Available at: https://peer.asee.org/improving-middle-school-girls-knowledge-self-efficacy-and-interests-in-sustainable-construction-engineering-through-a-steam-activated-program (Accessed July 07, 2022).

Osterman, K. F. (2000). Students' need for belonging in the school community. *Rev. Educ. Res.* 70, 323–367. doi: 10.3102/00346543070003323

Packard, B. (1999). "A 'Composite Mentor' Intervention for Women in Science." in *American Educational Research Association Annual Meeting; April 19–23,* Montreal, QC.

Phan, H. P. (2012). Relations between informational sources, self-efficacy and academic achievement: a developmental approach. *Educ. Psychol.* 32, 81–105. doi: 10.1080/01443410.2011.625612

Phan, H. P., and Ngu, B. H. (2016). Sources of self-efficacy in academic contexts: a longitudinal perspective. *Sch. Psychol. Q.* 31, 548–564. doi: 10.1037/spq0000151

Picciano, A. G. (2002). Beyond student perceptions: issues of interaction, presence, and performance in an online course. *JALN* 6, 21–40. doi: 10.24059/olj.v6i1.1870

Roxburgh, S. (2012). Parental time pressures and depression among married dual-earner parents. *J. Fam. Issues* 33, 1027–1053. doi: 10.1177/0192513X11425324

Ryan, R. M., and Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemp. Educ. Psychol.* 25, 54–67. doi: 10.1006/ceps.1999.1020

Sadker, M., and Sadker, D. (1994). *Failing at Fairness: How Our Schools Cheat Females*. New York: Simon & Schuster.

Sahin, A. (2013). STEM clubs and science fair competitions: effects on post-secondary matriculation. *J. STEM Educ.: Innov. Res.* 14, 5–11.

Salmons, J. (2015). *Qualitative Online Interviews: Strategies, Design, and Skills*. New York: Sage Publications.

Sax, L. J., Blaney, J. M., Lehman, K. J., Rodriguez, S. L., George, K. L., and Zavala, C. (2018). Sense of belonging in computing: the role of introductory courses for women and underrepresented minority students. *Soc. Sci.* 7:122. doi: 10.3390/socsci7080122

Scherpenzeel, A. (2002). Why use 11-point scales. *Swiss Household Panel (SHP)* 9:2008.

Scherpenzeel, A. C., and Saris, W. E. (1995). The quality of indicators of satisfaction across Europe. a meta-analysis of multitrait-multimethod studies. *A question of quality: evaluating survey questions by multitrait-multimethod studies*. Dissertation. Leidschendam: Royal PTT Nederland NV, KPN Research.

Seron, C., Silbey, S., Cech, E., and Rubineau, B. (2018). "I am not a feminist, but...": hegemony of a meritocratic ideology and the limits of critique among women in engineering. *Work. Occup.* 45, 131–167. doi: 10.1177/0730888418759774

Sevinc, B., Ozmen, H., and Yigit, N. (2011). Investigation of primary students' motivation levels towards science learning. *Sci. Educ. Int.* 22, 218–232.

Seymour, E. (1995). Guest comment: why undergraduates leave the sciences. *Am. J. Phys.* 63, 199–202. doi: 10.1119/1.17954

Seymour, E., and Hewitt, N. M. (1997). *Talking about leaving*. Boulder, CO: Westview Press.

Shapiro, C. A., and Sax, L. J. (2011). Major selection and persistence for women in STEM. *New Dir. Inst. Res.* 2011, 5–18. doi: 10.1002/ir.404

Simpkins, S. D., Davis-Kean, P. E., and Eccles, J. S. (2006). Math and science motivation: a longitudinal examination of the links between choices and beliefs. *Dev. Psychol.* 42, 70–83. doi: 10.1037/0012-1649.42.1.70

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

Sonnert, G., Fox, M. F., and Adkins, K. (2007). Undergraduate women in science and engineering: effects of faculty, fields, and institutions over time. *Soc. Sci. Q.* 88, 1333–1356. doi: 10.1111/j.1540-6237.2007.00505.x

Spear, M. G. (1984). The biasing influence of pupil sex in a science marking exercise. *Res. Sci. Technol. Educ.* 2, 55–60. doi: 10.1080/0263514840020107

St-Amand, J., Girard, S., and Smith, J. (2017). Sense of belonging at school: defining attributes, determinants, and sustaining strategies. *IAFOR J. Educ.* 5, 105–119. doi: 10.22492/ije.5.2.05

Stout, J. G., Dasgupta, N., Hunsinger, M., and McManus, M. A. (2011). STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *J. Pers. Soc. Psychol.* 100, 255–270. doi: 10.1037/a0021385

Strayhorn, T. L. (2018). *College students' sense of belonging: A key to Educational Success for All Students*. New York: Routledge.

Strenta, A. C., Elliott, R., Adair, R., Matier, M., and Scott, J. (1994). Choosing and leaving science in highly selective institutions. *Res. High. Educ.* 35, 513–547. doi: 10.1007/BF02497086

Sweeney, J. (2020). "The role of sub-discipline choice in women's enrollment and success within Canadian undergraduate engineering programs." *Proceedings of the Canadian Engineering Education Association (CEEA)*.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Rev. Educ. Res.* 45, 89–125. doi: 10.3102/00346543045001089

Tinto, V. (1987). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago, IL: University of Chicago Press.

Tinto, V. (1998). Colleges as communities: taking research on student persistence seriously. *Rev. High. Educ.* 21, 167–177.

Tinto, V. (2015). Through the eyes of students. *J. Coll. Stud. Retent.: Res. Theory Pract.* 19, 254–269. doi: 10.1177/1521025115621917

US Bureau of Labor Statistics (2021). Employment in STEM occupations. Available at: https://www.bls.gov/emp/tables/stem-employment.htm (Accessed July 07, 2022).

Usher, E. L., and Pajares, F. (2008). Sources of self-efficacy in school: critical review of the literature and future directions. *Rev. Educ. Res.* 78, 751–796. doi: 10.3102/0034654308321456

Vaz, S., Falkmer, M., Ciccarelli, M., Passmore, A., Parsons, R., Tan, T., et al. (2015). The personal and contextual contributors to school belongingness among primary school students. *PLoS One* 10:e0123353. doi: 10.1371/journal.pone.0123353

Walton, G. M., and Cohen, G. L. (2007). A question of belonging: race, social fit, and achievement. *J. Pers. Soc. Psychol.* 92, 82–96. doi: 10.1037/0022-3514.92.1.82

Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., and Zanna, M. P. (2015). Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering. *J. Educ. Psychol.* 107, 468–485. doi: 10.1037/a0037461

Wan, T. K. (2021). *Effects of extra-curricular project-based learning experiences on self-efficacy and interest in STEM fields in high school* (Doctoral dissertation, San Jose State University).

Wang, H. H. (2012). *A new era of science education: science teachers' perceptions and classroom practices of science, technology, engineering, and mathematics (STEM) integration*. University of Minnesota.

Wang, M. T., and Degol, J. (2013). Motivational pathways to STEM career choices: using expectancy–value perspective to understand individual and gender differences in STEM fields. *Dev. Rev.* 33, 304–340. doi: 10.1016/j.dr.2013.08.001

Wasburn, M. H., and Miller, S. G. (2004). Retaining undergraduate women in science, engineering, and technology: A survey of a student organization. *J. Coll. Stud. Retent.: Res. Theory Pract.* 6, 155–168. doi: 10.2190/NDXH-YM83-TKWY-4E6C

Wasburn, M., and Miller, S. (2005). "Can A Woman In Technology Freshman Seminar Change Student Attitudes? A Pilot Study" in *2005 Annual Conference*; June 12–15. 10.290.1–10.290.12.

Weaver, A. J., and Raptis, H. (2001). Gender differences in introductory atmospheric and oceanic science exams: multiple choice versus constructed response questions. *J. Sci. Educ. Technol.* 10, 115–126. doi: 10.1023/A:1009412929239

Webb, O., Wyness, L., and Cotton, D. (2017). *Enhancing Access, Retention, Attainment and Progression in Higher Education: A Review of the Literature Showing Demonstrable Impact*. London: Higher Education Academy.

Wilkinson, L. C., and Marrett, C. B. (1985). *Gender Influences in Classroom Interaction*. Cambridge, MA: Academic Press.

Wolffram, A., Derboven, W., and Winker, G. (2009). Women withdrawers in engineering studies: identity formation and learning culture as gendered barriers for persistence? *Equal Oppor. Int.* 28, 36–49. doi: 10.1108/02610150910933622

Wood, W., and Eagly, A. H. (2010). "Gender," in *Handbook of Social Psychology. 5th Edn. Vol. 1*. eds. S. T. Fiske, D. T. Gilbert and G. Lindzey (New York: John Wiley & Sons), 629–667.

Wood, W., and Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. *Adv. Exp. Soc. Psychol.* 46, 55–123. doi: 10.1016/B978-0-12-394281-4.00002-7

World Economic Forum (2021). These are the degrees that will earn you the most money when you graduate—and the ones that won't. Available at: https://www.weforum.org/agenda/2021/10/stem-degrees-most-valuable/ (Accessed July 07, 2022).

Zastavker, Y. V., Ong, M., and Page, L. (2006). "Women in engineering: exploring the effects of project-based learning in a first-year undergraduate engineering program" in *Proceedings. Frontiers in Education. 36th Annual Conference*; October 27–31; IEEE, 1–6.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemp. Educ. Psychol.* 25, 82–91. doi: 10.1006/ceps.1999.1016

Zohar, A., and Sela, D. (2003). Her physics, his physics: gender issues in Israeli advanced placement physics classes. *Int. J. Sci. Educ.* 25, 245–268. doi: 10.1080/09500690210126766

TYPE Original Research
PUBLISHED 29 August 2022
DOI 10.3389/fpsyg.2022.954325

Check for updates

OPEN ACCESS

EDITED BY
Milagros Sainz,
Open University of Catalonia, Spain

REVIEWED BY
Cristina Del Barrio,
Autonomous University of Madrid, Spain
Su Jiang,
Texas A&M University – Central Texas, United States

*CORRESPONDENCE
Kezia Olive
kezia.olive@helsinki.fi
Xin Tang
tangxin09@gmail.com

SPECIALTY SECTION
This article was submitted to Gender, Sex and Sexualities, a section of the journal Frontiers in Psychology

RECEIVED 27 May 2022
ACCEPTED 27 July 2022
PUBLISHED 29 August 2022

CITATION
Olive K, Tang X, Loukomies A, Juuti K and Salmela-Aro K (2022) Gendered difference in motivational profiles, achievement, and STEM aspiration of elementary school students. *Front. Psychol.* 13:954325. doi: 10.3389/fpsyg.2022.954325

Gendered difference in motivational profiles, achievement, and STEM aspiration of elementary school students

Kezia Olive*, Xin Tang*, Anni Loukomies, Kalle Juuti and Katariina Salmela-Aro

Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

To better understand the gender gap in science, technology, engineering and math (STEM) aspiration, the article examines the critical role of domain-specific motivation (i.e., expectancy and task values). Using longitudinal data from 5th and 6th grade (~11–12-year-old) students (n = 360, 55% girls), person-oriented analyses was applied to understand the gendered motivational profiles and their longitudinal influence on achievement and STEM aspiration. Specifically, we aimed to (1) derive motivational belief profiles regarding science, mathematics, and language (Finnish), (2) analyze the stability and change in the profiles between the 5th and 6th grade, (3) assess the relationship between motivational profiles and achievement and STEM aspiration, and (4) test for gender differences. We derived four motivational profiles for both years: high motivation in all subjects (~21%), high mathematics motivation (~46%), low mathematics motivation (~11%), and low motivation in all subjects (~8%). Latent transition analysis revealed that most students remained in the same profile throughout the 2 years. We found evidence of gendered differences in the motivational profiles and the chance of transitioning between profiles. More girls are characterized by low math motivation, while boys are more likely to transition to higher math motivation in 6th grade. The motivational difference is reflected in their achievement, although not strongly coupled with their STEM aspiration. The findings suggest that at this developmental stage, Finnish students have not developed a strong association between (gendered) STEM aspiration and their domain-specific motivation, although their motivation may have influenced their achievement. Interpretation and practical implications are discussed.

KEYWORDS

motivation, elementary school, expectancy-value theory, gender, STEM aspiration, latent profile analysis, latent transition analysis

Introduction

As part of the continuous effort to narrow the gender gap in science, technology, engineering, and mathematics (STEM), it is crucial to understand the factors that link gender and STEM involvement. Wang and Degol (2017) have discussed the various factors linking gender to differing levels of STEM engagement, and students' motivational beliefs are one of the most significant factors. Most scholars studying motivational beliefs and their relationship to gender and STEM involvement have been guided by situated expectancy value theory (SEVT) (Eccles et al., 1983; Eccles and Wigfield, 2020) and dimensional comparison theory (DCT) (Moller and Marsh, 2013). The SEVT framework provides a foundation for how personal characteristics, such as gender, influence students' motivational beliefs and academic outcomes. The DCT framework explains this process further, as it posits that each student will make internal comparisons between domains, which also significantly shapes their motivational beliefs (Moller and Marsh, 2013; Wigfield et al., 2020). Taken together, it is important to investigate motivational beliefs and their critical role in gendered STEM participation.

Yet, the relationship between gender, motivational beliefs, and STEM participation is less understood in elementary school students, even though the SEVT and DCT models indicate that motivational beliefs change over time (Eccles et al., 1993; Eccles and Wigfield, 1995; Guo et al., 2018b; Wan et al., 2021; Tang et al., 2022). Studies have found that students' beliefs about themselves and about different fields, such as STEM, are found to develop incrementally (Watson and McMahon, 2005; van Tuijl and van der Molen, 2015). During years prior to high school, students' interest and early educational experiences, especially in math and science, already sets the stage for their exploration and perceptions which predicts the subsequent choices they make (Hartung et al., 2005; Pinxten et al., 2012; Wang and Degol, 2017). These findings highlight the necessity of understanding the early years of students' motivational development and linking it to the factors influencing their outcomes.

Thus far, however, most studies that have addressed students' motivational beliefs and STEM aspiration have focused on data collected from adolescents (Gaspard et al., 2018, 2019; Guo et al., 2018b; Hsieh et al., 2019; Lazarides et al., 2021). Of the limited number of studies addressing elementary school students, researchers have examined only a few sets of domains and motivational constructs (Nurmi and Aunola, 2005; Gottfried et al., 2013; Musu-Gillette et al., 2015; Viljaranta et al., 2016; Petersen and Hyde, 2017). Consequently, there is a gap in our understanding of (gendered) motivational development given the limited number of studies involving elementary school students.

In this study, we aim to further understand the relationship between gender and STEM-related achievement and aspiration by examining the motivational beliefs of elementary school students in various subject domains. More specifically, we focus on students at the end of elementary school, right before they transition to (lower) secondary school, or junior high school. This will help to capture the transition period associated with significant changes in motivational beliefs (Watt, 2004) and offer new insights on the meaningful time frames in students' motivational belief development and its association with gender and important educational outcomes.

Theories on motivational beliefs

Situated expectancy value theory (SEVT) (Eccles et al., 1983; Eccles and Wigfield, 2020) focuses on an individual's motivational beliefs, processes of gender socialization, and choice behaviors. With this theory, motivational beliefs are conceptualized as a student's subjective task values (consisting of intrinsic value, utility value, attainment value, and cost) and expectations of success. *Intrinsic value* refers to the internal drive or enjoyment that a person has for a certain topic; *utility value* describes the future instrumental possibility of a certain behavior resulting in a particular goal (e.g., being good at math will help them in applying for an engineering degree); *attainment value* focuses on how a person attributes the importance of a certain behavior to their perceived identity (e.g., it is important for a girl like me to have good grades in languages); and *cost* refers to the perceived negative consequences for a person engaging with a certain task or behavior. In addition to addressing these values, the framework also described a student's expectations through *self-concept of ability,* which is an individual's perception of or belief about their ability level in a certain subject or domain (i.e., am I good at this? Can I really do this task?). The framework also described the influence of gender on an individual's hierarchy of values—the likelihood that an individual's personal characteristics, such as gender, will influence the hierarchy of their values, expectations, and choice behaviors (Eccles et al., 1993; Eccles, 2009).

Dimensional comparison theory (DCT) explains how these inner hierarchies develop across different domains or subjects in school (Moller and Marsh, 2013; Jansen et al., 2015). This theory proposes that students will compare their perceived performance in similar (e.g., math and science) and dissimilar domains (e.g., languages and math). They will then use the comparison to shape their motivational beliefs, including subjective task value hierarchies and expectancies in specific domains (Wigfield et al., 2020).

The combination of these theories has shown that there is a unique and rich process of motivational beliefs development in different domains within each individual. Both theories highlight the need to track the intra-individual hierarchy difference among students over time, a finding confirmed by different person-oriented studies (see a review by Wigfield and Eccles, 2020).

Motivational beliefs development

Another important assumption of the SEVT and DCT models is that students' motivational beliefs develop over time, though most studies to date have focused only on the motivational beliefs of adolescents (e.g., Guo et al., 2018b; Gaspard et al., 2019; Lazarides et al., 2021). The focus on an adolescent timeframe is understandable, as students become more and more stable in their differentiation of interests, confidence, and achievement in specific domains during those years (e.g., Gaspard et al., 2018, 2019; Lazarides et al., 2021; Wan et al., 2021). However, understanding the development of elementary school students is no less critical.

Elementary school students' motivational beliefs are an important foundation for further development, even if they are not as well differentiated as adolescents. As Eccles et al. (1993; Wigfield, 1994) have suggested, until the 5th grade, students have only developed a full understanding of the intrinsic value construct. They discovered that a full understanding of the other types of task values only start after this point. Nevertheless, studies conducted with students below 5th grade (which examined only their intrinsic values) still found domain-specific differences (Nurmi and Aunola, 2005; Viljaranta et al., 2016; Oppermann et al., 2021). Other studies with a longer time span and more task value dimensions have also confirmed persistent differences in intra-individual hierarchies of motivational beliefs starting from elementary school age (Archambault et al., 2010; Musu-Gillette et al., 2015). Furthermore, results from previous studies have also suggested that task value differences may begin in these early years and become the foundation for greater motivational gaps in older students (Guo et al., 2018b; Muenks et al., 2018).

Despite the suggestion that elementary school students' motivational beliefs are important foundation for further development of motivational belief, empirical evidence that clearly demonstrate this process is still scarce. Most longitudinal studies have only examined motivational beliefs in few domains independently, such as literacy (e.g., Archambault et al., 2010), mathematics (e.g., Musu-Gillette et al., 2015), or science (e.g., Vinni-Laakso et al., 2019). Even though several studies have considered more domains, they focused only on limited motivational constructs. For example, they mainly focused on the intrinsic motivation of students in the first years of schooling (Nurmi and Aunola, 2005; Oppermann et al., 2021). This highlights the need to provide more insight into the development of elementary school students' motivational beliefs in various domains and with respect to a comprehensive list of motivational belief constructs.

Motivational beliefs and academic outcomes

Different studies have confirmed that students' motivational profiles in different subjects indeed predict their academic outcomes, including achievement and STEM aspiration (for a review, see Wigfield et al., 2009). Scholars have found that differentiation of domain-specific motivation intensify during secondary school (Chow et al., 2012; Guo et al., 2018b; Lazarides et al., 2021; Wan et al., 2021) and predict students' subsequent achievement in the said domain (Bong et al., 2012; Safavian and Conley, 2016). The domain-specific mapping of motivation and achievement has also been observed in younger students (Nurmi and Aunola, 2005; Viljaranta et al., 2016).

Domain-specific motivational beliefs influence not only achievement but also students' future STEM career aspirations. Adolescent students who are characterized by a higher level of motivation in mathematics were more likely to aspire to a career in physical science and information technology (Chow et al., 2012; Guo et al., 2018b). Such students are also more likely to end up choosing a STEM-related career (Wang and Degol, 2013; Wang et al., 2015; Guo et al., 2018a). Studies with elementary school students have also yielded similar results, with higher science and math-related values predicting future STEM aspiration (Vinni-Laakso et al., 2019; Oppermann et al., 2021). This finding is in line with general findings showing that students' adult career choices (especially in STEM) are significantly influenced by their interests and self-concept as early as elementary school (Trice and McClellan, 1993; Maltese and Tai, 2009; van Tuijl and van der Molen, 2015; Lawson et al., 2018).

However, as mentioned before, most of the studies have focused on few subject domains and only certain aspects of motivational beliefs. Such a limitation means that more information is still needed especially with respect to understanding the domain-specific differentiation of motivational beliefs over time and its effects.

Gendered difference in motivational beliefs and academic outcomes

The relationships between motivational beliefs and academic outcomes, as assumed by the SEVT model, are also influenced by gender. Gendered differences in domain-specific motivation are evident from findings discussed in numerous studies, with boys being inclined more toward mathematics and girls toward languages, and they tend to remain the same for early elementary school students to adolescents in secondary school (Jacobs et al., 2002; Watt, 2008, 2016; Frenzel et al., 2010; Nagy et al., 2010). In one example from a person-oriented study, Gaspard et al. (2019) confirmed the overrepresentation of a low math motivational profile for girls, coupled with high motivation in languages. A similar finding has also been presented, for example, in studies by Eccles and Wang (2015), Umarji et al. (2018), Jansen et al. (2021), Lazarides et al. (2021), and Oppermann et al. (2021). These studies show that gendered motivational beliefs (with more girls being motivated to study languages and boys to study more math-intensive subjects) are linked to differing domain-specific achievement and aspiration or choice of university major for girls and boys.

More specifically, Guo et al. (2018b) assessed how gender influences students' personal *trajectories* with respect to the development of domain-specific motivational beliefs, which consequently shapes their occupational choices. Wang et al. (2013) also found that the main difference between the gender was influenced not only by the absolute levels of domain specific motivational beliefs but the different relative levels within the individual. Taken together, these studies stress the importance of intra-individual processes in the development of motivational beliefs, especially when considering the role of gender and its relation to academic outcomes.

To support younger students' STEM engagement, it is therefore also important to further identify how the dynamics of motivational belief, both at different development stages and in relation to gender, influence consequent achievement and aspiration. The unique intra-individual differences have also demonstrated the importance of accounting for insights from person-oriented approaches when identifying the sub-population differences in the development of gendered motivational beliefs.

The present study

In this study, we aim to extend current knowledge on the development of elementary school students' motivational beliefs and their role in achievement and STEM aspiration. We analyze data from Finnish 5th and 6th grade students (around 11–12 years old) to understand the development of motivational beliefs during the late elementary school years. We collected data on students' subjective task values and self-concept of ability in science, mathematics, and Finnish language to examine the effect of gender on the relationships involving motivational beliefs, achievement, and STEM aspiration. Specifically, we answer three research questions:

Research Question 1: (a) What motivational belief profiles can be identified from elementary school students in the domains of science, math, and Finnish language? (b) How stable are these profiles, and how likely are they to change from 5th to 6th grade?

We address the first question by analyzing and deriving motivational belief profiles for students in the 5th and 6th grades. Since other studies have already identified domain specific-profiles, such as math-specific and reading-specific profiles, in elementary school students (Nurmi and Aunola, 2005; Archambault et al., 2010; Viljaranta et al., 2016; Oppermann et al., 2021), we hypothesize that the domain-specific profiles are characterized by a clear differentiation in either science, math and/or Finnish. Additionally, following assumptions of DCT, we also hypothesize that levels of motivational beliefs should be similar in similar domains (i.e., science and math), and going opposite with dissimilar domains (i.e., math and/or science compared to language) (Hypothesis 1a). Moreover, since previous studies have also demonstrated the relative stability of these profiles, we also hypothesize that these domain-specific profiles are stable and consistent throughout 5th and 6th grade (Hypothesis 1b).

Research Question 2: To what extent are the motivational belief profiles associated with students' achievement and STEM aspiration?

Students' motivational profiles in science, math, and languages can predict academic outcomes with respect to achievement and STEM aspiration (Wigfield et al., 2009). Therefore, we expect to find a clear relationship between motivational belief in specific domains and achievement and STEM aspiration. Specifically, we assume that higher motivational belief in either math, science and/or Finnish will be reflected in higher achievement in the corresponding domain (Hypothesis 2a) and that higher motivational belief in math and science is also associated with higher STEM aspiration (Hypothesis 2b).

Research Question 3: To what extent do profiles, transitions, and their relation to students' STEM aspirations and achievement differ based on gender?

To provide further evidence on the influence of gender on students' motivation and academic outcomes, we focus on the relationship between the three. Previous studies based on the SEVT model have also discovered evidence of gendered differences in motivational profile membership, achievement, and STEM aspiration (Eccles, 2009; Wang and Degol, 2017; Wigfield and Eccles, 2020). Accordingly, we expect to find gender differences manifested in different ways. First, girls are overly represented in profile(s) identified with lower motivation in science and/or math and higher motivation in the Finnish language, and the profiles persist over time, while the profile memberships are the reverse for boys, but with similar persistence (Hypothesis 3a). Next, we also assume that girls—with lower science-math motivation levels—will have lower achievement scores in science and/or math, the opposite of boys (Hypothesis 3b). Girls with lower motivation in science-math will also have lower STEM aspirations, again the opposite of boys (Hypothesis 3c).

Materials and methods

Sample and procedure

Data from a Finnish longitudinal study (Name Removed for Reviewing Purpose) was used for this study, which followed students from seven schools in eastern Helsinki. The data collection process began in 2016 with first grade elementary school students at the age of seven or eight, and was always done each year in early February, which is in the middle of the school year. The development of their subject-specific motivation and aspirations were followed throughout the 6 years of elementary school.

In every data collection session, two researchers (or one assisted by a teacher) guided the students in answering the paper questionnaires. The researcher read each question and explained what each response means to the students (e.g., "one star in this one means I don't like science at all."). The assisting researcher or teacher walks around the class to check that every student can follow. During the data collection in year 5 and 6, either only one researcher is there or a teacher who have been trained administered the questionnaires as COVID pandemic situation limited the contact that is possible. The questionnaires were administered within one lesson (around 45 min) with short breaks in between.

The current study focused on students in their 5th and 6th grade (data collected in 2020 and 2021), and only students who participated in the data collection on those two waves are included. The final sample (N = 360, in 5th grade N girls = 200, boys = 160; in 6th grade, N girls = 192, boys = 164) had a mean age of 11.14 years (SD = 0.38) at 5th grade.

The study followed the ethical guidelines of the home institute. Parental consent was sought since participants in the study were elementary school-aged children. A description of the study and written permission forms were distributed to parents, and they had the opportunity to refuse to allow their child to participate in the study. Informed parental consent was obtained afterward for all the student participants. The headmasters and teachers from the participating schools were also informed about the study and agreed to the data collection schedule. Since the data collection was integrated with the students' normal classroom activities, the class teacher organized separate activities for students who did not have permission to participate in the study. Permission to collect students' data from schools was also obtained from the education division of the city of Helsinki (Kasko), with which we have cooperation agreements. According to the regulation from Kasko, no rewards or compensations are given for participants, either for students or the schools.

Finnish education context

The Finnish compulsory education system consists of mandatory schooling for children aged 7–18 years. Throughout grades 1 to 6, or the lower classes of the comprehensive school, the students had lessons in, among other subjects, mathematics, Finnish, and science—labeled "environmental studies"—which is a combination of biology, geography, physics, chemistry, and health education (Oppetushallitus, 2014).

Students in grades 3–6 received at least 2–3 h of science or environmental studies, mathematics, and Finnish language lessons per week, which accumulate to approximately 10–18 lessons throughout the 3 years[1]. In science, the lessons focused on students' knowledge and understanding, their research and working skills, and their values and attitudes toward the subject. In mathematics, the emphasis is on developing students' mathematical thinking to be logical, precise, and creative. In Finnish language, students' basic ability at listening, speaking, and reading is emphasized, while at the same time improving their self-expression, communication skills, and verbal awareness.

In terms of assessment, though students receive reports at the end of each school year, official national assessment criteria are only provided for students at the end of 6th grade to make sure the grades are comparable throughout the country. The grades ranged from from 4 to 10, with 5 as 'Pass' and 8 as 'Good.' The assessment criteria serves only as guidelines, and no national testing is conducted to determine different schooling tracks at this stage of schooling. Given this background, the Finnish context provided a unique opportunity to follow the development of elementary school students with less achievement-related feedback compared to some education systems in Europe (Hörner et al., 2015) such as in Germany or Austria, where students are streamed into different schooling tracks based on their achievements at the end of 4th grade.

Measures

At both measurement points (grades 5 and 6), we used the same student-reported subjective task value and ability self-concept questionnaire based on a scale developed by Eccles et al. (1993). This assessment was done for each of the three domains (mathematics, science, and Finnish language). Students were also asked about their dream occupation or aspirations in the form of an open-ended question. As this is a self-reported questionnaire, at each measurement point the teacher and/or a researcher and an assistant would instruct and assist students to make sure they understood the questions, scales, and responses expected of them.

Subjective task value

We assessed students' subjective task value with a Likert-type scale that ranged from 1 ("totally disagree") to 5 ("totally agree"). The response choices were shown as stars of an increasing number and size, following the 1–5 range.

1 https://www.oph.fi/sites/default/files/documents/distribution-of-lesson-hours-in-basic-education-2020.pdf

The students were asked to rate their intrinsic, utility, and attainment value in science, mathematics, and Finnish language domains. Intrinsic value was measured via three items: "I think the subject is fun"; "I like to do the schoolwork for this subject"; "I just like this subject." Utility value was measured via another three items: "Knowledge of this subject helps me during my free time"; "Knowledge of this subject will be useful for me in my future profession"; "The subject is useful for me." Finally, attainment value was also measured via three items: "I want to be good in (this subject)"; "I want to know a lot about this subject"; "This subject is important to me."

For the analysis, the average student score for intrinsic, utility, and attainment value was calculated to represent the subjective task value for each subject. All scales had good reliability at each measurement time and in all domains (Time 1: Science: $\alpha = 0.88$; Mathematics: $\alpha = 0.88$; Finnish language: $\alpha = 0.87$; Time 2: Science: $\alpha = 0.89$; Mathematics: $\alpha = 0.88$; Finnish language: $\alpha = 0.90$).

Ability self-concept

Self-concept of ability was assessed with three items following the same Likert-type visual response format: "I am good in (this subject)"; "I am good at the schoolwork for this subject"; "The schoolwork for this subject is easy to me" (1 star = "totally disagree," 5 stars = "totally agree").

Again, an average score for each subject was calculated. The reliabilities were as follows: Time 1: Science: $\alpha = 0.83$; Mathematics: $\alpha = 0.90$; Finnish language: $\alpha = 0.84$; Time 2: Science: $\alpha = 0.81$; Mathematics: $\alpha = 0.90$; Finnish language: $\alpha = 0.86$.

Achievement

Students' numerical grades for science, mathematics, and Finnish language were collected from the schools as a measure of student achievement. The grades are considered open information that is publicly accessible to all. Following Finnish school system mandates, a student is given a four as the lowest grade and a ten as the highest. On average, students in our sample have a mean of 8.414 ($SD = 1.011$) for science, 8.326 ($SD = 1.134$) for mathematics, and 8.354 ($SD = 1.021$) for Finnish language.

STEM aspiration

Students were asked an open-ended question about their dream jobs in both 5th and 6th grade, and their answers were combined to create a single aspiration variable and coded into occupational fields based on the ISCO-08 classifications (ILO, 2012). The encoding strictly followed the coding scheme. Based on these classifications, we derived two sets of coding schemes for the purpose of cross-validation. The *first coding scheme* included (a) mixed STEM fields (i.e., science and engineering professionals, health professionals, ICT professionals, science technicians, and associate professionals) and (b) non-STEM fields. The *second scheme* included (a) STEM-HBMS (health, biology, medical sciences), (b) STEM-MPCES (mathematics, physics, computer and engineering sciences), and (c) non-STEM fields.

Missing value analysis and outliers

The sample for this study ($N = 360$, grades 5 and 6) represents 51.4% of the initial 700 students who were part of the longitudinal sample followed from grade 1. The main reasons for the high attrition rate were because students changed schools before 6th grade or dropped out of the study due to the ethics permission renewal process. The data collection permit and parental consent had to be renewed by 2015, and in the process fewer parents responded and/or gave their consent, resulting in a significant degree of attrition.

The final sample of 360 students also include missing data ranging from 0.3% (Finnish subjective task value in grade 5) to 49% (STEM aspiration in grade 6). The exact percentages of missing values are shown in **Supplementary Table 1**.

A comparison of the missing values for each study variable showed that gender was related to missing values in STEM aspiration, with boys having significantly more missing values in both years ($p < 0.001$ for grade 5, $p = 0.009$ for grade 6). Moreover, students with missing values in aspiration and score in each subject generally had a significantly lower science and Finnish language self-concept at grade 5. A full comparison of missingness is shown in **Supplementary Table 5**.

We also checked for possible outliers with z $\geq/\leq$ 3.29, as suggested by Tabachnick et al. (2007), and found several potential multivariate outliers. These points were still included in the final analyses since they did not represent extreme values and did not affect the latent profile solutions.

Statistical analysis steps

Preliminary analysis

First, we checked the basic correlations and dependency between the variables, then conducted a confirmatory factor analysis for both self-concept of ability and subjective task values for each subject and measurement time.

This was followed by a measurement invariance test to confirm invariance assumptions about factor loadings, item intercept(s), and variance across the domains and two time points, and we found empirical support for strict measurement invariance. The model fit results were satisfactory for all steps, with CFI and TLI values being close to 0.95, SRMR values being close to 0.08, and RMSEA values being close to 0.06 (Hu and Bentler, 1999), and a decrease of less than 0.010 in CFI values and 0.015 in RMSEA values during every step, evidence of measurement invariance as recommended by Chen (2007).

Latent profile analysis

Next, we explored latent profile solutions for each measurement point separately, as suggested in previous longitudinal person-oriented study (Tang et al., 2021) and based on recommendations by Spurk et al. (2020). All models were estimated using Mplus 8.6 (Muthén and Muthén, 1998-2017) using a robust maximum likelihood estimator with the assistance of the R package MplusAutomation (Hallquist and Wiley, 2018).

We estimated up to six profiles using composite subjective task values and self-concept of ability for each subject (i.e., science, mathematics, and Finnish language) by freely estimating the means of these indicators. In terms of correlation and variances, we used the default model specification from Mplus: all variables are uncorrelated with all variables within the class and equal variances. To decide on the final number of profiles, we relied on both theoretical considerations by examining the difference in the mean for each profile and checking the fit information criteria. In this study, we relied on the Akaike Information Criterion (AIC), Consistent AIC (CAIC), Bayesian Information Criterion (BIC), and adjusted Bayesian Information Criterion (ABIC). Lower values for the four information criteria indicate a more optimal number of profile solutions. Visualization of the fit using elbow plots was used to compare the information criteria. The plot aided in deciding on the optimal solution by showing the number of profile solutions at which the slope started to flatten.

Latent transition analysis

After selecting the optimal number of profile solutions, we tested profile similarity by integrating the solutions from the two time points into a longitudinal latent profile analysis (LPA) model following the steps described by Morin and Litalien (2017). Four steps were followed: (1) *configural similarity* was tested to check if the numbers of profiles remained the same over time, using the same indicators with no constraints; (2) *structural similarity* was verified by constraining the indicator intercepts to note any similarities in global shape over time; (3) *dispersion similarity* was tested by constraining indicator variances over time to check the stability of within-profile variability; (4) *distributional similarity* was the final test, done by further constraining profile probabilities over time to confirm the stability of each profile's relative size. For each of these steps, two of the CAIC, BIC, and ABIC values should be lower compared to the last model to show evidence that the assumption is correct (Ryoo et al., 2018). After confirming the most similar model, we then converted it into a longitudinal latent transition analysis (LTA) model to identify stability and changes across latent profile membership over time.

Regression with predictors and outcomes

We used the final LTA model to test the extent to which students' profile membership and transition were related to their educational achievement and STEM aspiration. To examine the gender difference in profile membership, we used the three-step approach (R3STEP) in Mplus, as described by Asparouhov and Muthén (2014). We modeled gender as the predictor of latent profile membership through logistic regression. To evaluate gender difference in transition probabilities, we next used the KNOWNCLASS function. Finally, to test the extent to which gender and profile membership are related to students' outcomes, we applied a manual auxiliary three-step approach with a distal outcome (Asparouhov and Muthén, 2014). We used both gender and latent profiles as predictors, while treating the latent profiles as the auxiliary variable and regressing them based on students' grades (in science, math, and Finnish language) and STEM aspirations (with both coding schemes) as the outcome variable.

Results

Descriptive analysis

Means and zero-order correlations of the variables included in the analyses are reported in **Supplementary Table 1** for each motivational belief. The highest means for both self-concept and task values were in mathematics for students in both 5th and 6th grade. Mean comparisons showed a slight decrease in motivational beliefs in all subjects except for science self-concept. The results of the chi-square test of independence for the relationship between gender and STEM aspiration were significant, $\chi^2(1, N = 360) = 11.205$, $p < 0.00$, indicating a dependency between gender and aspiration (**Supplementary Table 2**). Measurement invariance testing for grades 5 and 6 showed that strict invariance of loadings, intercepts, and residual uniqueness was achieved for all three domain-specific self-concept and subject task value (STV) measures (see **Supplementary Table 3** for details of the fit summary).

Latent profiles of motivational belief and profile transition

The final profile solution was chosen based on theoretical meaningfulness, statistical criteria, and interpretability. The cross-sectional LPA for both time points suggested that the fit indices continued to improve as each profile was added, with lower AIC, CAIC, BIC, and ABIC values. However, the elbow plots at both time points showed that these fit indices started dropping less after the fourth profile, and profiles representing less than 5% of the participants emerged starting with the five-profile solution (**Supplementary Figure 1** and Supplementary Table 4). Therefore, after considering the fit indices, interpretation, and the meaningful distinction of the added profiles, we chose the four-profile solution for both time points.

Following this step, we employed a longitudinal LPA model to test similarities in the four-profile solution at both time points. Partial distributional similarity was retained with the lowest BIC and SABIC values. This implies that the number of profiles, intercept, variance, and group size were similar over time. We used this model for all further analyses. The profiles derived from this final model are illustrated in Figure 1.

With this four-profile solution, we labeled the first profile *high all* (grade 5 = 27.2%, n = 98; grade 6 = 15.3%, n = 55), as students in this profile showed high motivation in all domains. The second profile, with the most students, we labeled *high mathematics* (grade 5 = 51.4%, n = 185; grade 6 = 41.7%, n = 150), and it describes students with a moderate level of motivation in science and Finnish language and a high level of motivation in mathematics. In the third profile, *low mathematics* (grade 5 = 12.8%, n = 46; grade 6 = 9.4%, n = 34), students also exhibited a moderate level of motivation in science and Finnish language and low motivation in mathematics. The final and smallest profile, *low all* (grade 5 = 8.6%, n = 31; grade 6 = 7.2%, n = 26), describes students who reported low motivation in all subjects.

Following the four-profile solution, LTA provided the transition probabilities of each profile for students in the 5th and 6th grades. The probabilities are reported in Table 1. Students in the high *mathematics* profile exhibited the greatest stability (89%) followed by those in the low *mathematics* profile (78%). Some students also transitioned both to the math-specific profiles and to more general profiles. The highest rate of transition was observed for students moving from the *high all* to *high mathematics* profile (19%), followed by students moving from the *low mathematics* to *low all* profile, and vice versa (17% for each), and those moving from the *low all* to the *high all* profile (15%).

TABLE 1 Transition probabilities.

	Transition probabilities to 6th grade profiles			
Profiles at 5th grade	***High all***	***High math***	***Low math***	***Low all***
High all	**0.729**	0.191	0.074	0.006
High math	0.000	**0.894**	0.000	0.106
Low math	0.016	0.033	**0.781**	0.170
Low all	0.154	0.046	0.170	**0.630**
(Girls)				
High all	**0.746**	0.090	0.130	0.033
High math	0.000	**0.999**	0.000	0.001
Low math	0.026	0.001	**0.753**	0.220
Low all	0.140	0.047	0.271	**0.542**
(Boys)				
High all	**0.709**	0.268	0.024	0.000
High math	0.003	**0.790**	0.002	0.205
Low math	0.000	0.380	**0.620**	0.000
Low all	0.159	0.041	0.103	**0.696**

The values in bold represents profile stability from 5th to 6th grade.

Latent profile membership, achievement, and aspiration

With respect to grades, regression analyses found that students' membership in math-specific profiles (i.e., *high*

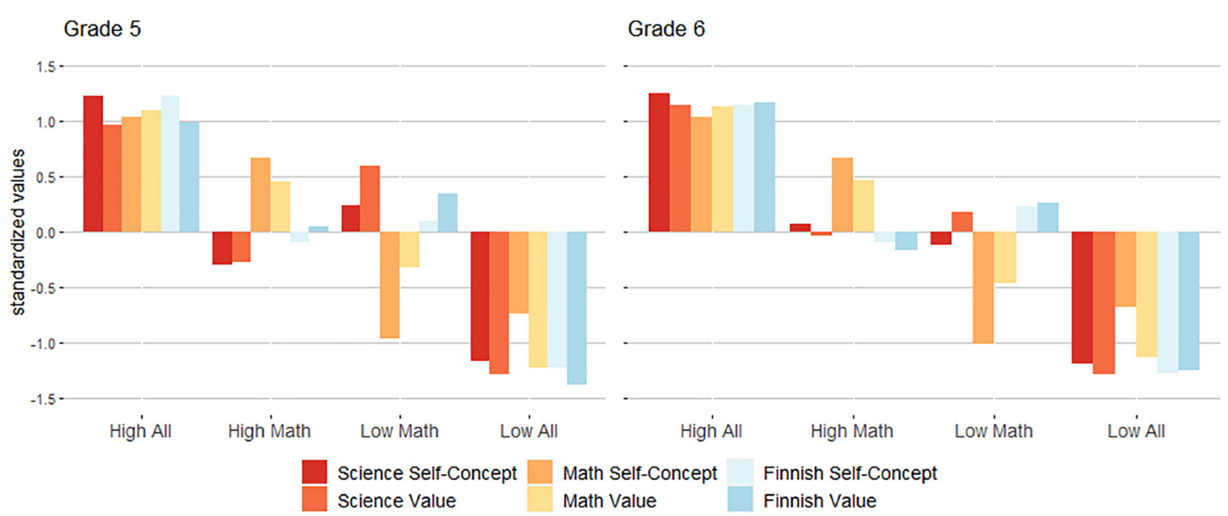

FIGURE 1
Profiles in both years.

TABLE 2 Achievement difference in each profile.

	High all (P1)	High math (P2)	Low math (P3)	Low all (P4)	Summary of significant differences
Science	9.016 [8.819;9.214]	8.653 [8.496; 8.811]	8.265 [8.035; 8.495]	7.784 [7.474; 8.094]	P1 > P2 > P3 > P4
Math	9.039 [8.854; 9.225]	8.828 [8.697; 8.960]	7.378 [7.062; 7.693]	7.613 [7.120; 8.106]	(P1 = P2) > (P3 = P4)
Finnish	8.929 [8.752; 9.105]	8.571 [8.373; 8.769]	8.309 [8.052; 8.567]	7.673 [7.253; 8.092]	P1 > (P2 = P3) > P4

Grades in Finnish schools are expressed in a 4-10 range; 4 as the lowest grade, 10 the highest.

TABLE 3 Science, technology, engineering and math (STEM) aspiration difference in each profile.

	High all (P1)	High math (P2)	Low math (P3)	Low all (P4)	Summary of significant differences
STEM Aspiration	0.406 [0.288; 0.523]	0.379 [0.271; 0.487]	0.254 [0.123; 0.384]	0.083 [0.029; 0.136]	P1 = P2 = P3 > P4
HBMS-MPCES Aspiration	0.566 [0.364; 0.769]	0.527 [0.349; 0.706]	0.297 [0.153; 0.440]	0.115 [0.054; 0.176]	P1 = P2 = P3 > P4

Aspiration was coded in two ways: 0 = Non-STEM, 1 = STEM; or 0 = Non-STEM, 1 = HBMS (Health, Bio and Medical Science), 2 = MPCES (Math, Physics, Computer and Engineering Sciences).

mathematics or *low mathematics*) are associated with differences in their levels of math achievement, as depicted in Table 2. Students in the *high mathematics* profile had similar math grades compared to students in the *high all* profile, although this finding cannot be generalized in the same way for science and Finnish language grades. We also detected a similar pattern when comparing students in the *low mathematics* and *low all* profiles, as they had similar math grades, but those in the former profile also had significantly higher science and Finnish language grades than students in the latter profile.

We also tested the association of the profiles with STEM aspiration, which revealed a different pattern compared to their achievement levels (Table 3). We found no STEM aspiration difference based on the math-specific profiles, unlike the achievement pattern, with only students in the *low all* profile showing significantly less STEM aspiration than those in all the other profiles. We noted no further difference between the other three profiles, as students in the *high all*, *high mathematics* and *low mathematics* profiles exhibited comparable levels of STEM aspiration. The patterns of the results are the same for both STEM aspiration coded for only STEM (mix) and those coded for HBMS and MPCES fields.

Gendered profile membership and transition

At both time points, we found gender differences in *low mathematics* profile, with more girls exhibiting moderate motivation and placed in this profile (78%; *n* = 36 in grade 5 and 79%; *n* = 27 in grade 6). This contrasts with the approximately equal distribution of boys and girls in all the other profiles (proportion of girls: ∼50% in *high all*, ∼54% in *high mathematics*, ∼56% in *low all*. See Supplementary Table 4). Logistic regression analysis showed that girls have a higher likelihood of being placed in this profile compared to other profiles, as described in Table 4.

In terms of transition, adding KNOWNCLASS to the model showed that girls especially exhibit higher levels of stability in the mathematic-specific profiles, as described in the lower part of Table 1. Girls have a 99% probability of remaining in the h*igh mathematics* profile (compared to 79% for boys) and a 75% probability of remaining in the lo*w mathematics* profile (compared to 62% for boys). Moreover, more boys seem to transition to the h*igh mathematics* profile (26% from h*igh all*, 38% from l*ow mathematics*, and 4% from l*ow all*) compared to girls, who exhibited a less than 10% transition probability from all the other profiles combined.

Gendered profiles, achievement, and aspiration

We found gendered differences in student academic performance at grade 6 within the different profiles. As described in Table 5, regression analysis revealed differences between girls and boys in the *high mathematics*, *low mathematics*, and *low all* profiles. With respect to students in the *high mathematics* profile, girls achieve significantly higher in science and Finnish language, but not in math. With respect to students in the *low mathematics* and *low all* profiles,

TABLE 4 Gendered difference in profile membership.

	Low math				High math				High all			
	B	*SE*	*p*	*OR*	*B*	*SE*	*p*	*OR*	*B*	*SE*	*P*	*OR*
Ref: Low all												
Female	−1.258	0.516	0.015	0.284	0.053	0.382	0.890	1.054	0.101	0.374	0.786	1.107
Ref: Low math												
Female					1.311	0.458	0.004	3.709	1.360	0.412	0.001	3.894
Ref: High math												
Female									0.049	0.295	0.869	1.050

TABLE 5 Gender effect on outcomes within profiles.

	High all			High math			Low math			Low all		
	B	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*
Achievement: Science	−0.143	0.232	0.536	**−0.437**	**0.192**	**0.023**	−0.367	0.255	0.151	−0.384	0.358	0.284
Achievement: Math	0.062	0.222	0.781	0.169	0.157	0.280	−0.515	0.315	0.102	−0.147	0.613	0.811
Achievement: Finnish	−0.287	0.203	0.157	**−0.680**	**0.201**	**0.001**	**−1.022**	**0.290**	**0.000**	**−0.951**	**0.447**	**0.033**
STEM aspiration	**−0.298**	**0.142**	**0.036**	0.015	0.138	0.915	−0.109	0.144	0.450	−0.028	0.060	0.640
HBMS-MPCES aspiration	−0.419	0.362	0.247	0.415	0.787	0.598	−0.231	0.183	0.208	−0.094	0.081	0.248

HBMS, Health, Bio and Medical Science; MPCES, Math, Physics, Computer and Engineering Sciences; Girls coded as 1; Boys 2. Significantly different profile are highlighted in bold.

girls also achieve significantly higher in Finnish language compared to boys.

In terms of STEM aspiration, presented in the lower part of **Table 5**, the differences between girls and boys within the profiles are not as visible. We found that girls in the *high all* profile have significantly more interest in aspiring to a STEM career. Otherwise, we detected no differences between girls and boys in terms of their STEM aspiration. Likewise, we found no difference when coding STEM for HBMS and MPCES. In other words, we observed no differences between girls and boys within each profile in terms of their aspiring to an HBMS, MPCES, or non-STEM occupation.

Discussion

Our study examines the influence of gender on motivational belief patterns and students' academic outcomes at the end of elementary school. Guided by expectancy value theory (Eccles et al., 1983; Eccles and Wigfield, 2020) and dimensional comparison theory (Moller and Marsh, 2013), we analyzed students' motivational patterns in specific domains and connected them to their levels of achievement and STEM aspirations. Our findings provide clear evidence of gender differences in students' motivational profiles, even among elementary school students, and the different profiles are associated with their achievement levels and aspirations.

Domain-specific motivational profiles and their stability

Our first aim was to identify intra-individual motivational patterns among students in the 5th and 6th grades. As a result of latent profile analysis, we identified four different motivational belief profiles throughout the 2 years (**Figure 1**). Two of the profiles were characterized by moderate motivation levels in science and Finnish language, one with a high motivation level in math (*high mathematics*), and the other profile with a low motivation level in math (*low mathematics*). The remaining two profiles showed a general pattern in all three domains (science, math, and Finnish language), one with high motivation levels for all domains (*high all*), and the other with low motivation levels (*low all*).

The first major finding partially supports our first hypothesis (1a) regarding a clear domain-specific differentiation in motivational belief profiles. The four profiles, two of which were characterized by strong motivation in math, confirmed that students have formed some domain-specific motivational beliefs already by 5th and 6th grade. This finding is similar to what had been reported in other person-oriented studies, such as studies by Nurmi and Aunola (2005), Viljaranta et al. (2016), and Oppermann et al. (2021). They also found that elementary school students have already developed clear intrinsic value and self-concept of ability toward mathematics. Our findings, however, add a new piece of evidence to the existing literature since we conceptualized motivational belief through task value perspective. By measuring subjective task value as a composite of intrinsic, attainment, and utility values (together with self-concept of ability), we still found that a clear domain-specific motivational profile for math has developed among students at this age.

On the other hand, we observed a lack of a specific motivational profile dedicated to science and/or language, contrary to the profiles found among older students. Past studies conducted using data collected from adolescents, such as those by Gaspard et al. (2018, 2019), Jansen et al. (2021), and Lazarides et al. (2021) identified profiles characterized by specific science and/or language motivation, not only specific profiles for mathematics. The difference between the results may be explained by the fact that the prior studies focused on older students, at which point students have developed more stable motivational profiles (Lazarides et al., 2016, 2019).

Additionally, the levels of motivational beliefs in the math-specific profiles suggested that the assumed *similar domains* (i.e., science and math) were growing in opposite directions, and what we assume as *dissimilar* (science and Finnish language) had similar levels of motivation. These profiles that we observed suggest that elementary school students have not developed the ability to distinguish the similarity and dissimilarity between the domains, particularly in science, as much as adolescents.

One of the reasons is that in the Finnish elementary school system the domain "science" is a mix of different subjects (i.e., biology, geography, physics, chemistry, and health education). Such a context most likely leads to less specialization in elementary school students since they perceive "science" less concretely. In other words, students at the Finnish elementary school stage have not been exposed to the differences between specific science domains (e.g., physics versus biology) or between the science domain and more language-intensive domains. After further exposure, older students could develop more domain-specific motivational profiles, as demonstrated by other person-oriented studies focusing on Finnish adolescents (Chow and Salmela-Aro, 2011; Guo et al., 2018b). We can, therefore, assume that further differentiation of domain-specific motivational profiles takes place later in students' development, as they become more exposed to the differences between domains, and not yet when they are in elementary school. This finding also resonates with the SEVT model (Eccles and Wigfield, 2020), which suggests that values are situationally bounded. A recent study also demonstrated that expectancy and task values are situative across domains, grade levels, and countries (Tang et al., 2022).

With regards to the transition between 5th and 6th grade, we found generally stable profile memberships, confirming hypothesis 1b, with only a few students shifting to different profiles. As suggested by the high odds of remaining in the same profile (above 60% for all profiles), students' general motivational level did not change during these years. This is true especially for students with high motivation. Some students did move to the more specialized *high mathematics* profile from the more general *high all* profile (around 19%), but generally they remained highly motivated. We noted a similar stable motivational trend for students with lower motivation, although with lower levels of stability.

The less stable profiles hint at the fact that students with lower motivation levels can still be pushed and encouraged to do better. Although 17% of students in either the *low all* or *low mathematics* profile remained in the low motivation profiles, 15% of students who were in the *low all* profile during 5th grade moved to the *high all* profile in 6th grade. This finding is encouraging in contrast to the more general trend that students will only continue to exhibit lower competence beliefs and task values as they grow older (Archambault et al., 2010; Musu-Gillette et al., 2015; Gaspard et al., 2017). During late elementary school years, some changes are still occurring and students also still develop an upward trajectory of motivation and not only a declining trajectory.

Domain-specific motivational profiles, achievement, and STEM aspiration

Our next major finding is that the domain-specific motivational profile is closely related to student achievement, as we expected for hypothesis 2a. We found that students fitting profile(s) with higher motivation levels in math tended to achieve better scores in the same domain. We also found the opposite effect to be true with respect to those students with low motivation levels in math. This finding is not as clearly demonstrated for the other two subjects, as we discovered no specific profiles identified with only a science or language focus. However, the general trend remained the same and confirmed our hypothesis: students in higher motivation profiles exhibited significantly higher achievement.

This finding confirms the direction of the relationship between domain-specific motivation and achievement that other studies have reported before. Even in elementary school students, the higher math motivation profile is associated with higher achievement or performance in this domain (Nurmi and Aunola, 2005; Viljaranta et al., 2016). This relationship most likely is a result of deeper engagement and persistent learning in domains where students already exhibit higher motivation (Wang et al., 2013; Tang et al., 2019).

In relation to aspiration, we found that the students' motivational profiles provide only limited information on their STEM aspiration. In this study, the *low all* profile was the only profile that successfully predicted lower STEM aspirations compared to the other profiles. Otherwise, we found no significant difference between all the other profiles in term of the students' STEM aspirations. More interestingly, although we identified profiles with a math-specific motivation, such motivation is not connected strongly to differences in STEM aspirations.

The limited association between domain-specific profile membership and STEM aspiration differs slightly from results presented in previous studies. Past studies have shown that early elementary school students characterized by higher math and science task values are more likely to have greater STEM aspirations (Vinni-Laakso et al., 2019; Oppermann et al., 2021). Studies done among older students have also found that those with higher math and science-specific motivation levels have greater aspiration to pursue a career in physical science or information technology (Guo et al., 2018b; Lazarides et al., 2021) and are the ones who typically end up choosing a STEM-related career (Wang and Degol, 2013; Wang et al., 2015; Guo et al., 2018a; Gaspard et al., 2019). In comparison, we did not find support for a strong association between membership in a math-specific profile and greater STEM aspirations.

The lack of evidence for such a strong association may suggest that students' STEM aspirations do not necessarily rely only on their math-specific motivation levels. A common finding for all the profiles showing comparable STEM aspiration (i.e., *high all*, *high mathematics*, *low mathematics*) is that they contained students with high and/or moderate science motivation levels. We can assume, therefore, that science motivation, in addition to math motivation, can also act as a source of interest contributing to strong STEM aspirations, even when it is not yet as well differentiated during elementary school. This assumption is also consistent when we consider the fact that students in the *low all* profile tend to have low motivation in all domains, leaving them with no buffer to even entertain the idea of aspiring to a STEM-related career path.

Taken together, even though elementary school students only show differentiation in terms of math-specific motivation levels, higher motivation in math and science is still associated with a greater likelihood of aspiring to a STEM-related career. With further differentiation, as described among older students, a clearer association between domain-specific motivation (in math and/or science) and STEM aspiration is more typically observed.

For elementary school students, we can only identify a clear distinction in math-specific profiles, even though we also considered other domains and all task value constructs. The profiles thus have only a limited relationship with STEM aspiration. This finding implies that strong coupling between clear subject-specific motivational beliefs and STEM aspiration has not taken place among elementary school students.

Gendered differences in motivational profiles

As we had expected with respect to hypothesis 3a, there are gendered differences in the motivational profile membership and the extent to which students transition between them. The logistic regression result showed that significantly more girls are in the *low mathematics* profile compared to other profiles. In terms of mathematics, we also discovered a difference in transition probabilities for both genders, with boys being more likely than girls to shift to the *high mathematics* profile in grade 6. These findings imply that indeed for students in their final years of elementary school, we can already observe gendered domain-specific motivational differences, especially in mathematics.

The gender differences among motivational profiles align with findings from previous studies. Person-oriented studies of older students have also shown that girls predominate in profiles characterized by a low level of motivation in mathematics (Chow and Salmela-Aro, 2011; Chow et al., 2012; Gaspard et al., 2019). This runs parallel with boys developing a higher math self-concept and greater self-confidence (Frenzel et al., 2010; Nagy et al., 2010). Longitudinally, Guo et al. (2018b) found that starting from grade 9, Finnish students tend to develop along different gendered motivational trajectories for different domains. Girls tend to place greater value on Finnish language and social subjects and less value on math and science. Exhibiting an opposite trend, boys place more value on math and science in their later school years.

The similar trend of motivational differences between girls and boys should serve as a warning of the risks to both genders. The transition odds for girls with low math motivation suggest that they will most likely continue to lose motivation for math, or even generally move more in the direction of lower general motivation. In other words, girls with low math motivation might be stuck in a vicious cycle of losing motivation in other subjects over time. This trend is not the same for boys: regardless of whether they have a higher or lower motivation in math in 5th or 6th grade, the odds are greater that they will end up in the higher motivational profiles in later years. It is important, therefore, to address the possibility that girls are at greater risk of continuing to lose motivation, especially in math, which will influence other outcomes as well.

Gendered motivational profiles and achievement

With respect to hypothesis (3b), we found that girls and boys achieved differently within certain profiles. In the *high mathematics* profile, the girls had significantly higher science and Finnish languages grades than the boys. The achievement gap also proved significant for the *low mathematics* and *low all* profiles, with girls having significantly higher Finnish language grades. Aside from the differences in science and Finnish language, we noted no gender differences in mathematic achievement within the profiles. In sum, it is worth noting that during these elementary school years, girls generally have higher achievement scores in science and Finnish language compared to boys.

It is interesting to note that we did not find evidence of gender difference in math grades within the profiles. This suggests that different math-specific motivation levels, as represented by the profiles, can sufficiently explain the differences in math achievement among students. Considering the fact that we also found gendered differences in the profile membership, this finding provides further evidence that gender influences students' outcomes through differences in math motivation. Consistent with SEVT, this result suggests that even among elementary school students, the influence of gender on domain-specific achievement is connected significantly with its influence on domain-specific motivation levels.

Moreover, with respect to DCT our findings indicate the possible start of a divergence in students' motivation levels and outcomes in 5th and 6th grades. According to the theory, one of the ways in which students are motivated to study a subject is through evaluating their performance in similar and dissimilar domains. Our results show that some girls who are equally as motivated in mathematics as boys still have higher achievement scores in science and Finnish languages. Based on the theory, the situation likely suggests that those girls will ultimately transition away from the math-intensive domain as they notice their strong performance in other dissimilar domains, such as in Finnish language.

Furthermore, in addition to processes related to domain-specific motivation, gender might also influence achievement through other means. Our results confirm that girls perform better in science and Finnish language, a finding confirmed by other studies as well (Wang and Degol, 2013; Keller et al., 2021), although we did not observe differences in the motivational profiles specifically for those domains. This indicates that gender also influences achievement through other processes beyond just domain-specific motivation, such as through different socialization processes (Eccles, 2009) and stereotypes (Miller et al., 2018; Master et al., 2021).

Gendered motivational profiles and science, technology, engineering and math aspiration

In contrast to hypothesis 3c, we did not find clear gender differences in STEM aspiration within domain-specific motivational profiles. The regression result showed no significant gender difference except for those within the *high all* profile, where more girls aspire to a STEM-related career. This result suggests that when the students are highly motivated in general, girls have higher STEM aspirations than boys. However, this gendered difference disappeared when we regressed STEM aspiration for HBMS and MPCES.

When taking this result into consideration along with the other findings, this study provides further understanding of gendered differences in association with motivation and STEM aspiration. The regression with gender provided evidence that only girls in the *high all* profile have significantly higher STEM aspirations. Based on these findings, we can assume that that low general motivation in elementary school students rather than higher levels of motivation makes a more significant difference in their STEM aspirations. However, if we are observing gender difference, it is only for the most part visible in more highly motivated students.

Another interesting point to note is that the result of the regression singled out highly motivated and high achieving girls as those having higher aspirations to pursue a STEM-related career. This result aligns with Finnish statistics of university students, which records that only around 25–30% of students enrolled are women in Information and Communication Technologies (ICT) or Engineering, Manufacturing and Construction, a contrast with Health and Welfare fields, in which 70% of students are women (StatisticsFinland, 2022). In other words, this finding seems to support the idea that high achieving girls, when they choose to enter STEM fields, are more likely to choose HBMS fields compared to MPCES.

On the other hand, this result is in contrast with our expectation that girls are the ones with less motivation to try hard in math and science, thus having less interest in pursuing a STEM-related career. The contradictory result in terms of more girls aspiring to a STEM-related career compared to boys may point to the tendency for girls to have a wider range of interests, higher levels of achievement, and therefore, more aspirational choice (Wang and Degol, 2013).

A note of caution is due here regarding our interpretations since the aspiration variable had the lowest response rate. A large proportion (49%) of the students did not answer the question about their dream job or they responded that they are unsure of their aspirations. In other words, a lack of awareness about possible career choices among students at this stage might be one possible reason for the lack of aspirational difference we observed.

Nevertheless, our findings still show that clear gendered differences with respect to higher STEM aspiration can only be detected among highly motivated students, which is slightly different from results presented in earlier studies. For instance, Guo et al. (2018b) found that throughout adolescence, girls have the tendency to exhibit decreasing levels of task values toward math and science, which is strongly associated with lower participation in STEM career. This finding also accords with a study by Oppermann et al. (2021) that focused on elementary school students' intrinsic value and self-concept in different domains. They found a similar result that stable math-specific motivation levels in elementary school students can mostly be observed in boys, and the motivation pattern was strongly associated with higher aspiration levels toward STEM.

The discrepancy between previous findings and our results may indicate that the development of students' STEM aspiration in late elementary school is not as straightforward. We should consider how much elementary school students understand and perceive different subject domains, whether similar or dissimilar ones, and different motivational constructs. According to our findings, most students did not have a well-developed and stable means of clearly differentiating between domain-specific motivation at this late elementary school stage, which is different than adolescents (Muenks et al., 2018; Wan et al., 2021). This developmental difference most likely explains the less clear associations between students' gendered motivation levels and outcomes, including their STEM aspirations. Moreover, as we argued in the previous section, students most likely have developed their aspiration at this point based on more information, such as certain stereotypes about which career is suitable for which gender (Chambers et al., 2018; Miller et al., 2018). However, this process might not yet necessarily manifest itself in the relationship mediated by motivation in 5^{th} and 6^{th} grade, and it is still likely that students from different profiles develop different STEM aspiration as they grow older (Mello, 2008; Lawson et al., 2018).

Implications of research

First, we found evidence that students' domain-specific differentiation takes place at the end of elementary school, even though it is not as well-differentiated. This finding supports the theoretical assumption regarding the developmental difference described by SEVT and the personal domain comparison processes described by DCT. Moreover, this finding also should support practices by educators. Understanding how students are developing different personal motivational beliefs for different domains at this stage should inform teachers' instructional processes.

Next, our findings also suggested the need to critically consider the possible trajectories for students' further motivational development and its impact on students' academic outcomes. We found an association between levels of math-specific motivation and achievement, which suggests the need for educators to pay specific attention to students' domain-specific motivation in supporting their achievement. Furthermore, we also observed less motivated students who transitioned to higher motivation profiles, hinting that there are still possibilities for change. Perhaps more possibilities exist for students to increase their motivational development in relation with achievement, which is a promising insight.

We also found evidence of gendered differences in the motivational profile membership, transitioning to higher motivation profiles, and academic outcomes. Significantly more girls in our study displayed low math motivation, while more boys transitioned to higher math motivation profiles in 6th grade. These gendered tendencies were significantly related to those particular students' achievement levels, but not to their STEM aspirations. Therefore, critical attention is needed to address the motivation levels and outcomes among students in both genders.

In terms of the gendered achievement gap, specific attention is needed for girls with lower motivation levels. Our results suggest that many girls already show low math motivation and a greater tendency to have even lower general motivation. Addressing this issue as early as possible is critical for such students, as it is more important for them not to continue dropping in their motivation and achievement levels, thus preventing the continuation of a vicious cycle.

On the other hand, our results also suggest critical points to be addressed further in terms of enhancing aspiration and interest in STEM. The evidence we found suggests that in this age group, gendered motivation is still quite malleable and is not reflected clearly in students' aspiration levels. For instance, we found girls who still *have* a more open attitude toward math continue to perform well in this domain, and girls who have equal, if not higher, academic performance in all domains compared to boys, even when they have similar levels of motivation. Some of the more highly motivated girls even showed greater interest in STEM. However, previous studies also found that they are also the students most likely to be steered away from such choices as their value hierarchy is increasingly influenced by their broader achievements and interests (Wang et al., 2013).

It is therefore important to to develop further studies addressing the dynamics of factors related to why girls end up not pursuing STEM careers, especially in MPCES, despite their high motivation and achievement. Especially in the Finnish context – understanding the dynamics between early subjective STEM experiences and social and/or environmental barriers is necessary (Schoon, 2001).

Furthermore, it is also critical to help students stay motivated and encourage them to consider a STEM-related field for future studies and/or work through designing interventions for students at this stage. For example, as we found that students seem to not base their aspiration on their domain-specific interests, development of interventions aimed at exposing them to examples and possibilities to not only cultivate their STEM interests but also STEM career aspiration (e.g., from exposing them to different role models and narratives as suggested by Luttenberger et al. (2019). Additionally, as our findings also hints at potential effects of stereotypes and socialization effects in STEM interest development, interventions designed to challenge STEM-stereotypes that students might have developed is necessary. Most importantly, activities and programs that support students' STEM interest in terms of understanding a wider STEM relevance at school (e.g., as described by Harackiewicz et al., 2014; Gaspard et al., 2015) should be of prime importance to maintain students' STEM motivation.

Limitations and further research

Our study provides further insight into the relationship between gender, motivational beliefs, and their longitudinal effect on students' achievement and STEM aspirations at the end of elementary school. However, certain limitations need to be considered. First, in terms of statistical power, our sample was not large enough to provide further details on different associations. For example, with the current sample size we could not test the relationship between the transition within 2 years and the students' academic outcomes. We also could not officially test the association between the profiles with specific STEM aspiration (HBMS-MPCES) since our sample was not large enough compared to the very limited response rate from students.

Second, we only used data from two time points (in Grade 5 and 6) to assess the students' motivational development, with achievement data only for grade 6. These limitations do not allow a complete insight into students' overall motivational development in elementary school, especially in relationship with a key outcome, such as achievement. Thus, future studies should consider a longer time span to provide more nuanced understanding of students' longitudinal motivation development and its relationship with key outcomes.

Finally, we derived most variables from self-report measures, except for the students' grades. Further studies that also focus on the influence of different factors in shaping students' academic outcomes would need to consider more sources of information.

Conclusion

This study provides evidence that students in the last years of elementary school have developed different motivational profiles associated with their academic outcomes. We identified four different motivational profiles, two of which were specific to math but none specific to science or the verbal domain. This domain specialization remains stable throughout the school years and is most likely only enhanced as students grow older. We identified an association between higher math motivation levels and higher grades in each of the profiles. On the other hand, math-specific profile membership was not strongly connected to higher STEM aspiration, although general low motivation is associated with lower STEM aspiration. Taken together, our findings provide evidence that domain comparison processes are indeed already underway even among elementary school students, and that different profile membership influences students' achievement and aspiration in different ways.

We also provided more insight on the relationship between critical outcomes, such as students' grades and STEM aspiration, and their motivational beliefs and gender. Girls and boys showed different tendencies for profile membership and transition. In general, girls are showing significantly lower motivation in math and lower transition toward higher motivation in math. This gendered tendency was clearly reflected in their outcome, such as their math achievement scores. On the other hand, girls showed higher achievement in science and Finnish compared to boys with similar motivation, and some girls with high motivation even showed higher STEM aspiration. These findings present the different opportunities and risks for their development that requires further exploration.

In sum, to support academic outcomes for both girls and boys it is important to consider their gendered motivational beliefs. Understanding these associations is important in light of supporting students' development along different career pathways, and future studies can and should build upon these findings to further identify critical periods and constructs for intervention and improvement.

Data availability statement

The data analyzed in this study was subject to the following licenses/restrictions: this study used an already existing dataset that includes sensitive data (pseudonymized dataset linked to personal identity of students). The raw dataset will only be shared publicly according to the regulations from Academy of Finland scheme by the PIs. Queries regarding these datasets should be directed to the corresponding authors KO, kezia.olive@helsinki.fi or XT, tangxin09@gmail.com.

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 to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

KO performed the statistical analysis together with XT. KO wrote the manuscript with suggestions and evaluations from XT and KS-A. XT and KS-A co-supervised the whole project. KJ and AL collected the data and provided extensive context information. All authors discussed the results and contributed to the final manuscript.

Funding

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 953326. The research was supported by the Academy of Finland grants #298323 and #336138, awarded to KS–A.

Acknowledgments

We kindly thank the different members of our research group: Johannes Gale, who worked as research assistant and contributed to coding parts of the dataset and managing the data; Dr. Junlin Yu, who gave advice and assistance in interpreting the results and statistical modeling alongside the R mates. We also thank all the teachers and schools participating in the Helsinki area for making it possible to collect the data even during the pandemic.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.954325/full#supplementary-material

References

Archambault, I., Eccles, J. S., and Vida, M. N. (2010). Ability self-concepts and subjective value in literacy: Joint trajectories from grades 1 through 12. *J. Educ. Psychol.* 102, 804–816. doi: 10.1037/a0021075

Asparouhov, T., and Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using mplus. *Struct. Equat. Mod. Multid. J.* 21, 329–341. doi: 10.1080/10705511.2014.915181

Bong, M., Cho, C., Ahn, H. S., and Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *J. Educ. Res.* 105, 336–352. doi: 10.1080/00220671.2011.627401

Chambers, N., Kashefpakdel, E., Rehill, J., and Percy, C. (2018). *Drawing the Future: Exploring The Career Aspirations Of Primary School Children From Around The World*. London: Education and Employers.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equat. Mod. Multid. J.* 14, 464–504. doi: 10.1080/10705510701301834

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.

Chow, A., Eccles, J. S., and Salmela-Aro, K. (2012). Task value profiles across subjects and aspirations to physical and IT-related sciences in the United States and Finland. *Dev. Psychol.* 48, 1612–1628. doi: 10.1037/a0030194

Eccles, J. S. (2009). Who am i and what am i going to do with my life? Personal and collective identities as motivators of action. *Educ. Psychol.* 44, 78–89. doi: 10.1080/00461520902832368

Eccles, J. S., Adler, T. E., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). "Expectancies, values, and academic behavior," in *Achievement and Achievement Motives*, ed. J. T. Spence (San Fransisco: W. H. Freeman and Company), 75–138.

Eccles, J. S., and Wang, M.-T. (2015). What motivates females and males to pursue careers in mathematics and science? *Int. J. Behav. Dev.* 40, 100–106. doi: 10.1177/0165025415616201

Eccles, J. S., and Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personali. 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

Eccles, J. S., Wigfield, A., Harold, R. D., and Blumenfeld, P. (1993). Age and gender differences in child. *Child Dev.*64, 830–847. doi: 10.1111/j.1467-8624.1993.tb02946.x

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. Adoles.* 20, 507–537. doi: 10.1111/j.1532-7795.2010.00645.x

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B., Häfner, I., Nagengast, B., et al. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Dev. Psychol.* 51:28. doi: 10.1037/dev0000028

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., and Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemp. Educ. Psychol.* 48, 67–84. doi: 10.1016/j.cedpsych.2016.09.003

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., and Marsh, H. W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemp. Educ. Psychol.* 52, 1–14. doi: 10.1016/j.cedpsych.2017.10.003

Gaspard, H., Wille, E., Wormington, S. V., and Hulleman, C. S. (2019). How are upper secondary school students' expectancy-value profiles associated with achievement and university STEM major? A cross-domain comparison. *Contemp. Educ. Psychol.* 58, 149–162. doi: 10.1016/j.cedpsych.2019.02.005

Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., and Oliver, P. H. (2013). Longitudinal pathways from math intrinsic motivation and achievement to math course accomplishments and educational attainment. *J. Res. Educ. Effect.* 6, 68–92. doi: 10.1080/19345747.2012.698376

Guo, J., Eccles, J. S., Sortheix, F. M., and Salmela-Aro, K. (2018a). Gendered pathways toward STEM careers: The incremental roles of work value profiles above academic task values. *Front. Psychol.* 9:1111. doi: 10.3389/fpsyg.2018.01111

Guo, J., Wang, M.-T., Ketonen, E. E., Eccles, J. S., and Salmela-Aro, K. (2018b). Joint trajectories of task value in multiple subject domains: From both variable- and pattern-centered perspectives. *Contemp. Educ. Psychol.* 55, 139–154. doi: 10.1016/j.cedpsych.2018.10.004

Hallquist, M. N., and Wiley, J. F. (2018). MPlus automation: An R package for facilitating large-scale latent variable analyses in mplus. *Struct. Equat. Mod.* 2018, 1–18. doi: 10.1080/10705511.2017.1402334

Harackiewicz, J. M., Tibbetts, Y., Canning, E., and Hyde, J. S. (2014). Harnessing values to promote motivation in education. *Adv. Motiv. Achiev.* 18, 71–105. doi: 10.1108/S0749-742320140000018002

Hartung, P. J., Porfeli, E. J., and Vondracek, F. W. (2005). Child vocational development: A review and reconsideration. *J. Vocat. Behav.* 66, 385–419. doi: 10.1016/j.jvb.2004.05.006

Hörner, W., Döbert, H., Reuter, L. R., and Von Kopp, B. (2015). *The education systems of Europe*. Dordrecht: Springer.

Hsieh, T., Liu, Y., and Simpkins, S. D. (2019). Changes in united states latino/a high school students' science motivational beliefs: Within group differences across science subjects, gender, immigrant status, and perceived support. *Front. Psychol.* 10:380. doi: 10.3389/fpsyg.2019.00380

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equat. Mod. Multid. J.* 6, 1–55. doi: 10.1080/10705519909540118

ILO (2012). *International standard classification of occupations 2008 (ISCO-08): Structure, group definitions and correspondence tables*. Geneva: International Labour Office.

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., Becker, M., and Neumann, M. (2021). Dimensional comparison effects on (gendered) educational choices. *J. Educ. Psychol.* 113:330.

Jansen, M., Schroeders, U., Lüdtke, O., and Marsh, H. W. (2015). Contrast and assimilation effects of dimensional comparisons in five subjects: An extension of the I/E model. *J. Educ. Psychol.* 107, 1086–1101. doi: 10.1037/edu0000021

Keller, L., Preckel, F., Eccles, J. S., and Brunner, M. (2021). Top-performing math students in 82 countries: An integrative data analysis of gender differences in achievement, achievement profiles, and achievement motivation. *J. Educ. Psychol.* 2021:685. doi: 10.1037/edu0000685

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

Lazarides, R., Dicke, A. L., Rubach, C., Oppermann, E., and Eccles, J. S. (2021). Motivational profiles across domains and academic choices within Eccles et al.'s situated expectancy-value theoretical framework. *Dev. Psychol.* 57, 1893–1909. doi: 10.1037/dev0001250

Lazarides, R., Dietrich, J., and Taskinen, P. H. (2019). Stability and change in students' motivational profiles in mathematics classrooms: The role of perceived teaching. *Teach. Teacher Educ.* 79, 164–175. doi: 10.1016/j.tate.2018.12.016

Lazarides, R., Viljaranta, J., Aunola, K., Pesu, L., and Nurmi, J.-E. (2016). The role of parental expectations and students' motivational profiles for educational aspirations. *Learn. Indiv. Diff.* 51, 29–36. doi: 10.1016/j.lindif.2016.08.024

Luttenberger, S., Paechter, M., and Ertl, B. (2019). Self-concept and support experienced in school as key variables for the motivation of women enrolled in STEM subjects with a low and moderate proportion of females. *Front. Psychol.* 10:1242. doi: 10.3389/fpsyg.2019.01242

Maltese, A. V., and Tai, R. H. (2009). Eyeballs in the fridge: Sources of early interest in science. *Int. J. Sci. Educ.* 32, 669–685. doi: 10.1080/09500690902792385

Master, A., Meltzoff, A. N., and Cheryan, S. (2021). Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proc. Natl. Acad. Sci. U.S.A.* 118:48. doi: 10.1073/pnas.2100030118

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

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

Moller, J., and Marsh, H. W. (2013). Dimensional comparison theory. *Psychol. Rev.* 120, 544–560. doi: 10.1037/a0032459

Morin, A. J., and Litalien, D. (2017). *Webnote: Longitudinal tests of profile similarity and latent transition analyses*. Montreal, QC: Substantive Methodological Synergy Research Laboratory.

Muenks, K., Wigfield, A., and Eccles, J. S. (2018). I can do this! The development and calibration of children's expectations for success and competence beliefs. *Dev. Rev.* 48, 24–39. doi: 10.1016/j.dr.2018.04.001

Musu-Gillette, L. E., Wigfield, A., Harring, J. R., and Eccles, J. S. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educ. Res. Evalu.* 21, 343–370. doi: 10.1080/13803611.2015.1057161

Muthén, L. K., and Muthén, B. O. (1998-2017). *Mplus User Guide*. 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. Adoles.* 20, 482–506. doi: 10.1111/j.1532-7795.2010.00644.x

Nurmi, J.-E., and Aunola, K. (2005). Task-motivation during the first school years: A person-oriented approach to longitudinal data. *Learn. Instruct.* 15, 103–122. doi: 10.1016/j.learninstruc.2005.04.009

Oppermann, E., Vinni-Laakso, J., Juuti, K., Loukomies, A., and Salmela-Aro, K. (2021). Elementary school students' motivational profiles across Finnish language, mathematics and science: Longitudinal trajectories, gender differences and STEM aspirations. *Contemp. Educ. Psychol.* 64:101927. doi: 10.1016/j.cedpsych.2020.101927

Oppetushallitus (2014). *National core curriculum for basic education 2014*. Helsinki: Finnish National Board of Education.

Petersen, J. L., and Hyde, J. S. (2017). Trajectories of self-perceived math ability, utility value and interest across middle school as predictors of high school math performance. *Educ. Psychol.* 37, 438–456. doi: 10.1080/01443410.2015.1076765

Pinxten, M., De Fraine, B., Van Den Noortgate, W., Van Damme, J., and Anumendem, D. (2012). Educational choice in secondary school in flanders: the relative impact of occupational interests on option choice. *Educ. Res. Eval.* 18, 541–569. doi: 10.1080/13803611.2012.702991

Ryoo, J. H., Wang, C., Swearer, S. M., Hull, M., and Shi, D. (2018). Longitudinal model building using latent transition analysis: An example using school bullying data. *Front. Psychol.* 9:675. doi: 10.3389/fpsyg.2018.00675

Safavian, N., and Conley, A. (2016). Expectancy-value beliefs of early-adolescent hispanic and non-hispanic youth. *AERA Open* 2:357. doi: 10.1177/2332858416673357

Schoon, I. (2001). Teenage job aspirations and career attainment in adulthood: A 17-year follow-up study of teenagers who aspired to become scientists, health professionals, or engineers. *Int. J. Behav. Dev.* 25, 124–132.

Spurk, D., Hirschi, A., Wang, M., Valero, D., and Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *J. Vocat. Behav.* 120:103445. doi: 10.1016/j.jvb.2020.103445

StatisticsFinland (2022). *University students and degrees (National classification of education 2016) by year, university, gender, field of education, level of education and information*. Helsinki: Statistics Finland.

Tabachnick, B. G., Fidell, L. S., and Ullman, J. B. (2007). *Using multivariate statistics*. Boston, MA: Pearson.

Tang, X., Lee, H. R., Wan, S., Gaspard, H., and Salmela-Aro, K. (2022). Situating expectancies and subjective task values across grade levels, domains, and countries: A network approach. *AERA Open* doi: 10.1177/23328584221117168

Tang, X., Wang, M. T., Parada, F., and Salmela-Aro, K. (2021). Putting the goal back into grit: Academic goal commitment, grit, and academic achievement. *J. Youth Adoles.* 50, 470–484. doi: 10.1007/s10964-020-01348-1

Tang, X., Wang, M.-T., Guo, J., and Salmela-Aro, K. (2019). Building Grit: The longitudinal pathways between mindset, commitment, grit, and academic outcomes. *J. Youth Adoles.* 48, 850–863. doi: 10.1007/s10964-019-00998-0

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

Umarji, O., McPartlan, P., and Eccles, J. S. (2018). Patterns of math and English self-concepts as motivation for college major selection. *Contemp. Educ. Psychol.* 53, 146–158. doi: 10.1016/j.cedpsych.2018.03.004

van Tuijl, C., and van der Molen, J. H. W. (2015). Study choice and career development in STEM fields: An overview and integration of the research. *Int. J. Technol. Des. Educ.* 26, 159–183. doi: 10.1007/s10798-015-9308-1

Viljaranta, J., Aunola, K., and Hirvonen, R. (2016). Motivation and academic performance among first-graders: A person-oriented approach. *Learn. Indiv. Diff.* 49, 366–372. doi: 10.1016/j.lindif.2016.06.002

Vinni-Laakso, J., Guo, J., Juuti, K., Loukomies, A., Lavonen, J., and Salmela-Aro, K. (2019). The relations of science task values, self-concept of ability, and STEM aspirations among finnish students from first to second grade. *Front. Psychol.* 10:1449. doi: 10.3389/fpsyg.2019.01449

Wan, S., Lauermann, F., Bailey, D. H., and Eccles, J. S. (2021). When do students begin to think that one has to be either a "math person" or a "language person"? A meta-analytic review. *Psychol. Bull.* 147, 867–889. doi: 10.1037/bul0000340

Wang, M. T., and Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy-value perspective to understand individual and gender differences in STEM fields. *Dev. Rev.* 33:1. doi: 10.1016/j.dr.2013.08.001

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

Wang, 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

Wang, M.-T., Degol, J., and Ye, F. (2015). Math achievement is important, but task values are critical, too: Examining the intellectual and motivational factors leading to gender disparities in STEM careers. *Front. Psychol.* 6:36. doi: 10.3389/fpsyg.2015.00036

Watson, M., and McMahon, M. (2005). Children's career development: A research review from a learning perspective. *J. Vocat. Behav.* 67, 119–132. doi: 10.1016/j.jvb.2004.08.011

Watt, H. M. (2016). "Gender and motivation," in *Handbook of motivation at school*, eds D. B. Miele and K. R. Wentzel (New York, NY: Routledge), 320–339.

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. (2008). A latent growth curve modeling approach using an accelerated longitudinal design: The ontogeny of boys' and girls' talent perceptions and intrinsic values through adolescence. *Educ. Res. Evalu.* 14, 287–304. doi: 10.1080/13803610802249316

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educ. Psychol. Rev.* 6, 49–78. doi: 10.1007/BF02209024

Wigfield, A., and Eccles, J. S. (2020). "Chapter Five - 35 years of research on students' subjective task values and motivation: A look back and a look forward," in *Advances in Motivation Science*, ed. A. J. Elliot (Elsevier), 161–198.

Wigfield, A., Eccles, J. S., and Möller, J. (2020). How dimensional comparisons help to understand linkages between expectancies, values, performance, and choice. *Educ. Psychol. Rev.* 32, 657–680. doi: 10.1007/s10648-020-09524-2

Wigfield, A., Tonks, S., and Klauda, S. L. (2009). *"Expectancy-Value Theory," Handbook of Motivation at School*. New York, NY: Routledge/Taylor & Francis Group, 55–75.

TYPE Original Research
PUBLISHED 09 September 2022
DOI 10.3389/fpsyg.2022.944377

OPEN ACCESS

EDITED BY
Katja Upadyaya,
University of Helsinki, Finland

REVIEWED BY
Milagros Sainz,
Open University of Catalonia, Spain
Susana González-Pérez,
CEU San Pablo University, Spain

*CORRESPONDENCE
Joanna Pyrkosz-Pacyna
jpyrkosz@agh.edu.pl

SPECIALTY SECTION
This article was submitted to Gender, Sex and Sexualities, a section of the journal Frontiers in Psychology

RECEIVED 15 May 2022
ACCEPTED 05 August 2022
PUBLISHED 09 September 2022

CITATION
Pyrkosz-Pacyna J, Dukala K and Kosakowska-Berezecka N (2022) Perception of work in the IT sector among men and women—A comparison between IT students and IT professionals. *Front. Psychol.* 13:944377. doi: 10.3389/fpsyg.2022.944377

Perception of work in the IT sector among men and women—A comparison between IT students and IT professionals

Joanna Pyrkosz-Pacyna[1]*, Karolina Dukala[2] and Natasza Kosakowska-Berezecka[3]

[1]Faculty of Humanities, AGH University of Science and Technology, Kraków, Poland, [2]Faculty of Psychology in Katowice, SWPS University of Social Sciences and Humanities, Katowice, Poland, [3]Faculty of Social Sciences, University of Gdansk, Gdańsk, Poland

Lack of gender balance within STEM fields is caused by many complex factors, some of which are related to the fact that women do not perceive certain occupations as congruent with their career and personal goals. Although there is a large body of research regarding women in STEM, there is a gap concerning perception of occupations within different STEM industries. IT is a domain where skilled employees are constantly in demand. Even though the overall female representation in STEM fields is rising and that the IT industry is undertaking numerous interventions to attract women to careers in IT, the representation of women in this domain is still disappointingly low. Therefore, the goal of our study was to examine the possible differences among male and female IT and non-IT students and employees in terms of their perception of IT and other key factors influencing the feeling of aptness of IT as a potential sector one's career: goal congruence, sense of belonging and self-efficacy. In this paper we present the results of a study conducted in Poland among working IT professionals ($N = 205$) and IT students ($N = 127$) that we compare with individuals from non-IT sectors ($N = 222$ employees, 107 students). Our results showed significant gender differences between IT students and IT professionals. We found that communal goals are more important for IT employees than for IT students (both male and female) and that a sense of social belonging is stronger among female IT employees than among male IT employees and IT students. Women employed in IT also had the same level of sense of social belonging as women in non-IT group. These findings suggest that after entering IT positions, women's perception of the domain might become potentially more favorable and attuned with their needs. We also found that female IT students value agentic goals more than communal goals which was not the case for female IT employees. The results highlight the importance of investigating women's perception of the IT sector at different levels of career in terms of their goals and other work-related variables. Such lines of research will help develop more effective interventions in attracting women to enter the IT field.

KEYWORDS
STEM, IT, gender, work interests, work perception

Introduction

Despite numerous interventions aimed at reaching gender equality in STEM (Science, Technology, Engineering, Mathematics) domains women are still vastly underrepresented in many STEM-related professions. STEM represents multiple and very distinct technological disciplines among which some are more, and some are less represented by women. Among the industries most highly dominated by men is IT (Information Technology). The proportion of women in IT in USA is reported to be ∼25% (Ashcraft et al., 2016; Fry et al., 2021), with more detailed data from 2020 showing that women make up 28–42% of the GAFAM (Google, Apple, Facebook, Amazon, and Microsoft; Statista, 2021). Nevertheless these women who enter IT jobs, are less likely than men to work as programmers and tend to occupy positions of a tester or project manager. According to the Women in IT (2022) 70% of IT female specialists reported difficulties in entering this line of work. Similarly to the labor market, within academia unequal representation of women and men is visible. Among all STEM majors, IT manifests one of the largest gender gaps. In domains such as mathematics, physics, or life-sciences the female to male students' ratio is less unbalanced (Fry et al., 2021). Yet in IT the change in female representation over the years is barely visible (Jasko et al., 2020; European Commission Directorate-General for Research and Innovation, 2021; Fry et al., 2021). For example, the Women in Polytechnics Report (Knapińska, 2022) shows that the largest ever increase in the number of female IT students in Poland recorded the historical spike from 12% in 2014 to 14% in year 2018. Authors of this report argue that this spike in admissions might stem from multiple initiatives and interventions aimed at women in IT such as "IT for SHE," "Geek Girls Carrots," "Girls in Tech" or "Women in Technology". This is however not an informed conclusion as to our best knowledge, no data is available proving effectiveness of programs in enhancing women's intention to enter and remain in IT sector.

There are however some results available showing the effectiveness of interventions aimed at more equal gender representation in STEM. For example, Dennehy and Dasgupta (2017) showed in the longitudinal experiment, that having a female peer mentor early in college increases women's positive academic experiences and retention in engineering. In the study by Ramsey et al. (2013) interventions focused on creating a welcoming academic environment resulted in decreased stereotyping concerns and increased implicit STEM identification among STEM female students. Interestingly, there is also evidence for unsuccessful interventions. Cowgill et al. (2021) for example demonstrated how programs overtly emphasizing women's minority status in STEM might lead to opposite effects such as decrease in women's interest in STEM. It is important to mention at this point, that all the above results pertain to interventions conducted with the US samples. The data from other regions is scarce. For example in a study by Peña et al. (2021) focusing on diagnosing the extent to which gender issues were even mentioned within STEM teaching programs, 84% of the teaching staff stated that their proposed activities did not include any gender aspects. It is also worth noting that these scarce interventions that are rarely evaluated usually pertain to STEM as a discipline in general—to our knowledge there are no domain specific analyses that would focus on IT only.

Previous literature provided explanations for possible reasons for low female representation in STEM among university majors and among employees (for example: Dweck, 2007; Cheryan et al., 2009; Diekman et al., 2010; Singh et al., 2013). Less is known about women particularly in the IT field. It seems to be a unique domain where on the one hand women are constantly on the margins and on the other hand where the demand for highly qualified staff is continuously rising (United Nations Technology and Innovation Labs Report, 2019; Fry et al., 2021). What is more, most of the research undertaking the topic of women's lower representation in STEM has been focused on western samples, with fewer studies conducted within the context of Eastern and Central Europe. Our study thus contributes to current research lines focused on lower female representation in the IT sector by adding another under-researched context that of Poland.

Drawing on the well-established theoretical approaches (i.e., Goal Congruence Theory, Self-Efficacy and Sense of Social Belonging) we carried out an empirical research exploring the factors most relevant to exclusion of girls and women from STEM, specifically the field of IT. Our research contributes to finding solutions for low representation of women in the highly male-dominated sector of IT by focusing on perception of this domain among men and women. We have also looked at different stages of the career path within IT by including both IT students and IT employees in our sample. Specifically we examine the possible differences among male and female IT and non-IT students and employees in terms of their perception of IT and other key factors potentially influencing the feeling of aptness of IT: goal congruence, sense of belonging and self-efficacy. We examine women's perceptions of IT while they are IT students and when they are IT professionals. Comparing women at two different career stages allows us to verify how the potential personal misfit to IT may be detrimental for women who comply with commonly accepted social norms. Such lines of research help define areas in which women are mostly vulnerable to gender stereotypes and thus decide to quit this career.

One of the factors contributing largely to low female representation in STEM domains is the lack of perceived goal congruence. Women tend to prioritize communal goals and therefore might discard IT as a domain that does not afford those goals – as a result they don't feel they belong to the IT sector. Additionally, women in STEM experience self-efficacy decline along their career progress (Brainard and Carlin, 1998;

Sterling et al., 2020; Stewart et al., 2020) which, when combined with the conviction that IT requires high intellectual capabilities, discourages them from entering this career field. This may suggest why women, even when skilled in other scientific matters, opt out from choosing a career in IT as early as when choosing their study majors.

Women in IT sector: Goal congruence

In recent years researchers provided valuable insights into possible mechanisms contributing to low female representation in STEM, which can be crucial when analyzing the situation of women in the IT sector. The obtained results referred to in the literature highlight the importance of analyzing how women perceive a given sector as allowing them to realize their career goals. The Goal Congruence Theory (Diekman et al., 2010; Brown et al., 2015; Steinberg and Diekman, 2017) focuses on two categories of people's goals: communal (i.e. working with others, helping others or cooperation) and agentic (i.e. power, status, career development or expertise growth). It tackles two aspects of these goals: goal endorsement (the type of goals that are personally important and are targeted for pursuit) and goal affordance (perception of possibility to achieve certain goals) both of which may influence the decision to engage in certain domains. In a series of studies, Diekman et al. (2010) showed that STEM occupations are perceived as affording communal goals to a lesser extent than other jobs and that this perception among women in turn influences the decisions regarding future career orientation. Since women tend to lean toward occupations that afford communal goals congruence, the perception of STEM might be one of the key reasons for low female representation in those domains. These results are especially valid for the IT industry as it might be perceived as a technical and solitary type of occupation. Numerous studies point out the existence of specific IT specialist stereotypes that entail isolation, social inadequacy, over-focus on technology, unpopularity, and most importantly, masculinity (Herz, 1997; Sanger et al., 1997; Clayton et al., 2009; Santos et al., 2022). This might deter women choosing IT as their career path.

On the other hand, analyses of possible indicators of high female drop-out ratio show the importance of not only communal goals endorsement but also the need to meet their agentic goals. An increasing number of studies show that agentic goals are becoming equally important for men as well as for women (Moore et al., 2008; Pyrkosz-Pacyna et al., 2019). Yet satisfying agentic goals (in women's case) in IT is not as evident as one may expect. Fouad et al. (2012) found that reduced access to key creative roles, and a sense of feeling stalled in one's career are among others the most significant factors contributing to female attrition from the tech field. Additionally, women who left IT positions were less likely to report opportunities for training and development and support from a manager within their former workplace. It would therefore seem that even though IT is perceived as an agentic sector, women in IT are uncertain if they can attain their agentic goals. In turn, lack of goal congruence might impact another important variable relating to women's intention to stay within the IT sector both when they are students and when they are professionals—that variable is the sense of social belonging.

Sense of social belonging and women's representation in the IT sector

The sense of social belonging is the conviction that one is accurately fitting into a given environment. It also entails perception of social connectedness in groups and the sense of fitting in socially with others (Baumeister and Leary, 1995; Walton and Cohen, 2007). Numerous studies have explored the importance of social belonging in reference to women's representation in STEM. These studies (Cheryan and Plaut, 2010; Good et al., 2012; Tellhed et al., 2017; Aelenei et al., 2020) showed that indeed women tend to feel that they do not naturally belong in STEM and conversely, by experimentally increasing their sense of fit, women declared more interest in career in STEM. Very subtle cues can make the shift away or toward STEM. For example, in a study by Cheryan et al. (2009) altering the surroundings of college career advisory meetings (reflecting stereotypical tech-geek space: sci-fi posters, video games stations or empty energy drink cans vs. neutral one) influenced the women's willingness to consider IT as a college major. Even women already working in IT notice the specific image of this industry, admitting that the "geekiness" is a part of the IT image (Moore et al., 2008).

At the same time, there is evidence that despite the possible lowered sense of belonging women feel the connection with the IT field. The study conducted at Technological University (Pyrkosz-Pacyna et al., 2019) revealed that women's experience of a sense of belonging to the University was the same as that of men. The existing evidence is scarce and mixed. It requires further analyses that would allow us to explore the level of women's sense of belonging to the IT sector at different career levels.

Women's self-efficacy in the IT sector

According to the value - expectancy theory (Eccles, 1987) people tend to engage in endeavors that are viewed as both valuable and reachable. The classic study by Bandura (1997) showed that self-efficacy, that can be defined as person's belief in his or her capability to successfully perform a particular task, is related to the quantity of required effort and the willingness to persist at tasks. If the goal is perceived as outside one's competences, the action toward it might be weak or forsaken altogether. The perception of one's own capabilities plays an important role in considering various career choices. Numerous studies showed that women in general tend to have lower self-efficacy especially in STEM fields, even when controlling for the actual educational outcomes (Correll, 2001; Singh et al., 2007; Good et al., 2012; Sterling et al., 2020). Additionally, science is perceived as a domain where those with exceptional innate talents are more frequently represented (Dweck, 2007). Women also tend to hold themselves to much higher standards when it comes to what they perceive as high skills (Correll, 2004), and they report lower self-efficacy than men regardless of their actual performance (Kost-Smith et al., 2009; Stout et al., 2011). As a result, their self-efficacy convictions are strongly related to engagement in STEM. Studies showed that low math self-efficacy predicts low intention to pursue a STEM career (Correll, 2001) and similarly the computer self-efficacy (Miura, 1987, 2020). It is important to highlight that self-efficacy is not necessarily representative of one's actual skills. There is robust evidence showing that there are no observable gender differences in STEM competences (U.S. Department of Education, 2007) yet the self-perception of those skills is continuously biased, with men assessing their science skills more positively than women (Correll, 2001; Stewart et al., 2020). The bias against women in the IT sector also plays a role here. The stereotypes that women are less skilled in programming tasks can cause great damage. Studies by Terrell et al. (2016) showed that codes prepared by men and by women were assessed equally, but only if the coder gender was not revealed. Yet, as in the case of previous factors, further studies are needed to explore the perception of women's skills in the IT sector, which in turn has a significant impact on their engagement in this domain.

In our research we focus on two groups of women: (1) women who have chosen IT as their major and (2) those who are employed in IT companies. We compare those results to the results of men and to comparable samples of women studying or working in non-IT sectors. This allows us to gather data about women at different stages of their IT career trajectory. The transition from university to employment is a unique time period. Data shows that even women who graduated from IT majors are likely to forgo their career in this field (Gu, 2018) which is a huge waste of their previous efforts and their unique vocational potential. Our sample consisting of middle and eastern European participants is also uniquely valuable as Poland is a country with considerable IT outsourcing facilities. The market for IT employees in Poland is significant (Kossowska et al., 2012). Results of our study contribute to the knowledge about factors significant in designing and implementing various interventions for women in STEM with specific focus on the IT field.

The present research

To explore possible differences between perception of future work conditions and the actual experience of employees in the IT sector and inform future interventions, our sample consists of (male and female) IT students and IT employees. To make sure that these findings are specific to the IT field, we have included a comparison group composed of non-IT women, including management students and banking employees (male and female) that do not belong to STEM fields. The obtained results can contribute to designing effective interventions both specifically in central European setting and in general by showing how work in IT is perceived in multiple perspectives: that of men and women, on different levels of their career trajectory and in comparison to other domains. The knowledge of various factors impacting attitudes toward work in IT can be applied in designing interventions based on familiarity with the targeted group of interest. For example, perhaps different interventions might be suitable for women in IT jobs than for IT students when the intervention is tackling sense of social belonging to the domain. Detailed hypotheses are described below.

Goal endorsement and goal affordance hypotheses

We predicted that women in IT will value communal goals more than men in IT (H 1a) but less than women working in non-IT fields (H 1b). Since IT is perceived as mostly affording agentic goals, we suspected those women who decided to go into IT will value these goals more than women in non-IT sample (H 1c). We also overall predicted that women in IT will value communal goals more than agentic goals (H 1d). In our study we also wanted to test whether there are any differences between female IT employees and IT students. These analyses were exploratory, along with between-field, IT vs non-IT comparisons.

When it comes to perception of IT in terms of goal affordance, we predicted that women in IT will perceive lower goal affordance, both communal (H 2a) and agentic (H 2b) than men, and then women in non-IT working environments (H 2c and H 2d respectively). As before we also explored differences between female IT employees and students and across fields (IT vs. non-IT).

Social belonging hypotheses

Since IT is dominated by men and is stereotypically considered as a manly profession, we predicted that women in IT will declare a lower sense of social belonging than men in IT (H 3a) and non-IT women (H 3b). Again, we explored the differences between female IT employees and students and across IT and non-IT fields.

Self-efficacy hypotheses

We predicted, according to existing literature, that women in IT will have lower self-efficacy than men in IT (H 4a) and then non-IT professional women due to perceiving IT as a domain where special skills usually connected to masculinity are needed (H 4b). We also explored the differences among female IT employees and students regarding their self-efficacy. Again, we explored the differences across fields.

Method

Procedure

The study was conducted in early 2019 (before COVID-19 pandemic), among IT (informatics) majors and management students and among IT and management professionals. IT and management students were contacted via email with a link to the survey attached. For taking part in the study, participants were offered the possibility to participate in a draw to receive a gratuity in the form of a university gadgets gift box (containing fountain pen or USB drive). For the sake of anonymity participants were informed that to take part in the draw they should provide their contact information in a separate link available at the end of the survey, so it will be impossible to link their results with personal information. Participants could fill in the questionnaires only in an online form, in their free time.

Professionals were recruited from two companies at their Polish divisions: an international banking corporation (with an expanded IT department) and an IT company. We wanted to include a sample from non-STEM and at the same time not stereotypically female industry and banking/management is a good example of such industry. The invitation to take part in the study was sent out by the internal communication system to both IT and non-IT departments of the companies. Participants were not offered any remuneration for taking part in the study. Since one of the companies where we conducted the study is international and hires international staff, upon the request of the company representatives, the survey was also available in English for non-Polish participants. However, due to the very small sample size, we have excluded answers of English speaking participants from further analysis. Participants filled in online questionnaires while at work.

Participants

A total sample of 724 participants took part in the study. 63 surveys were dropped because participants did not indicate their gender or were outside the IT/banking sector, so 661 surveys were used for further analysis. Final sample consisted of 127 IT students (73 female), 107 control group students (57 female), 205 IT employees (82 female) and 222 control group employees (137 female). Mean age of employees was $M = 31.04$ ($SD = 6.08$) and $M = 22.64$ ($SD = 3.59$) of students. 180 students were on the bachelor's degree level and 54 were master's degree students.

In the sample of employees mean seniority in the company was M = 3.27, SD = 3.13, minimum 0 years and maximum 26 years.

Measures

All scales were back translated into Polish by a native speaker. All items were adjusted for employees/students participants (e.g., "Do you feel confident about your line of work?" for employees and "Do you feel confident about your line of studies?" for students).

Goal congruence

Scale developed by Diekman et al. (2010). The scale consists of 23 goals in two dimensions: 14 on agency (e.g., achievement, power) and 9 on communality (e.g., helping others, serving humanity). Using a scale from 1 to 7, participants were asked to determine how much each goal is personally important for them (goal endorsement) and how much their work domain (or major in the case of students) allows them to achieve those goals (goal affordance), where 1 = unimportant/ this domain does not enable achieving this goal, and 7 = very important/ this domain enables achieving this goal. Reliability of the scale was Cronbach's $\alpha = 0.840$ for agency endorsement, Cronbach's $\alpha = 859$, communal endorsement, Cronbach's $\alpha = 0.916$ for agency affordance and Cronbach's $\alpha = 0.836$ for communal affordance (all coefficients were computed for the whole sample, both employees and students).

Self-efficacy

Scale developed by Dennehy and Dasgupta (2017), consist of 6 items (e.g., Do you think you have a talent for your line of work? Do you feel confident about your line of work?) on 1 (I disagree) to 7 (I agree) Likert scale (Cronbach's $\alpha = 0.703$).

Sense of social belonging

Scale developed by Dennehy and Dasgupta (2017) consist of four statements, e.g., "I feel connected to my colleagues in my field" and measures an individual's sense of belonging to the given field, with 1 to 7 scale (1 = I disagree, 7 = I agree) (Cronbach's $\alpha = 0.826$).

Analysis strategy

The goal of our analysis was to show the specificity of the situation of women in the IT field, therefore we compared women's results to (1) men working in the IT sector and to (2) women working outside of the IT sector. We later compared IT female employees with IT students' results. Hence, analyses of all dependent variables are mostly conducted in three steps: in step 1 we compare male and female employees in the IT field (MANOVA with gender as an IV, we present those analyses in Table 1). In step 2 we compare female employees and students in the IT vs. non-IT field (MANOVAs with area of work as an IV, these analyses are presented in Table 2). In the last step we compare both IT female employees and IT female students and IT male employees to IT male students (MANOVAs with work status as IV, all those analyses are presented in Table 3). Other analyses are mentioned in the text. Other information as well as the syntax and database can be found in the supplementary materials. All correlations between variables are in Table 4. Means, confidence intervals and the differences between means in all groups are in Table 5.

Results

Goal endorsement

First, we tested if women in IT will value communal goals more than men in IT (Hypothesis 1a). We observed a significant difference [$F_{(1,203)} = 6.48$, $p = 0.012$, $\eta^2 = 0.03$]: female employees in IT valued communal goals more than male. We observed a lack of effect for female students in IT compared to male students [$F_{(1,127)} = 3.66$, $p = 0.058$, $\eta^2 = 0.03$].

Contrary to our hypothesis 1b, female employees in IT rated communal goal endorsement at the same level as women outside of IT [$F_{(1,217)}) = 0.02$, $p = 0.880$, $\eta^2 < 0.01$]. A different pattern was observed for female students [$F_{(1,128)} = 9.14$, $p = 0.003$, $\eta^2 = 0.07$]: they perceived communal goal endorsement at a lower level than women studying outside of IT.

Similarly, male employees in IT rated communal goals at the same level as men in the comparison group [$F_{(1,206)} = 0.19$, $p = 0.664$, $\eta^2 = 0.01$], and male students in IT perceived communal goals at a lower level than students outside IT [$F_{(1,102)} = 7.17$, $p = 0.009$, $\eta^2 = 0.07$].

In the next step we were interested in perception of agentic goals. We investigated the difference between male and female IT employees, and we found no difference in terms of agentic goals for female and male employees [$F_{(1,203)} = 0.01$, $p = 0.905$, $\eta^2 = 0.01$] and students [$F_{(1,125)} = 0.06$, $p = 0.799$, $\eta^2 = 0.01$].

Next, we checked the difference between women employees in IT vs. non-IT in terms of perception of agentic goals (Hypothesis 1c). We observed no significant difference [$F_{(1,217)} = 1.52$, $p = 0.219$, $\eta^2 = 0.01$], the same lack of effect was for female students [$F_{(1,128)} = 3.46$, $p = 0.065$, $\eta 2 = 0.03$]. There were also no differences for male IT vs. non-IT employees in terms of importance of agentic goals [$F_{(1,206)} = 0.01$, $p = 0.986$, $\eta^2 = 0.01$], but male IT students had lower agentic goal endorsement than male students outside IT [$F_{(1,102)} = 6.80$, $p = 0.011$, $\eta^2 = 0.06$].

Finally, we tested if women in IT value communal goals more than agentic goals (Hypothesis 1d). We conducted repeated measures ANOVA only for women in IT, with the importance of communal and agentic goals as dependent variables. The difference between the importance of the two types of goals was not significant [$F_{(1,81)} = 0.53$, $p = 0.469$, $\eta^2 < 0.01$]. Female employees in IT value agentic and communal goals at the same level as female employees outside of IT. The pattern was different for IT female students. Female students in IT value agentic goals more than communal goals [$F_{(1,72)} = 10.47$, $p = 0.002$, $\eta^2 = 0.127$]. On the other hand, men in IT rated agentic goals higher than communal goals, which was true for employees [$F_{(1,122)} = 18.38$, $p < 0.001$, $\eta^2 = 0.13$] and students [$F_{(1,53)} = 26.17$, $p < 0.001$, $\eta^2 = 0.35$].

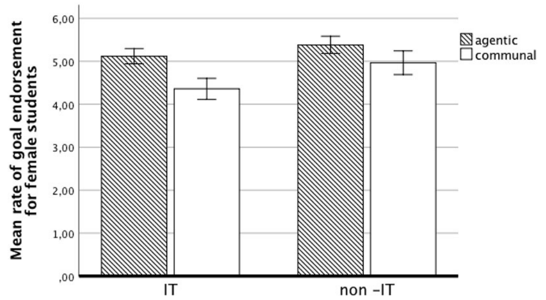

FIGURE 1
Mean rate of goal endorsement for female students.

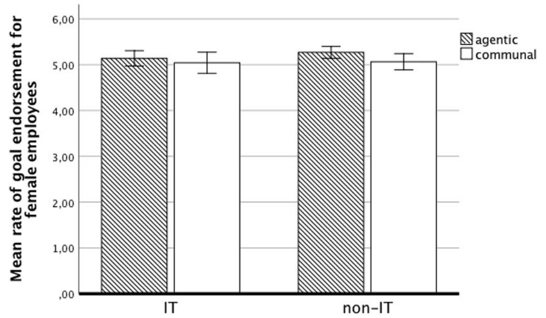

FIGURE 2
Mean rate of goal endorsement for female employees.

In the last step we compared female IT students and female IT employees in terms of goal endorsement. The difference was not significant for agentic goals [F $_{(1,153)} = 0.03$, $p = 0.858$, $\eta^2 = 0.01$], but was significant for communal goals [$F_{(1,153)} = 16.13$, $p < 0.001$, $\eta^2 = 0.10$]: female IT employees had higher communal goal endorsement than did the students. Differences in goal endorsement between women in IT and non-IT field (both students and employees) are presented in Figures 1, 2.

The same pattern was for observed for men: male IT employees rated agentic goals at the same level as male students [$F_{(1,175)} = 0.11$, $p = 0.741$, $\eta^2 = 0.01$] however, they had higher communal goal endorsement than did the students [$F_{(1,175)} = 12.59$, $p < 0.001$, $\eta^2 = 0.07$].

Goal affordance

First, we compare women in IT to men in IT in terms of goal affordance (Hypothesis 2a and 2b). MANOVA with communal goal affordance as a dependent variable showed no significant differences between female and male employees [$F_{(1,203)} = 2.92$, $p = 0.089$, $\eta^2 = 0.01$] and students [$F_{(1,127)} = 0.22$, $p = 0.693$, $\eta^2 = 0.01$]. The difference for agentic goal affordance as a dependent variable was not significant between male and female IT employees [$F_{(1,203)} = 1.48$, $p = 0.225$, $\eta^2 = 0.01$] but was significant for male and female IT students [$F_{(1,127)} = 9.15$, $p = 0.003$, $\eta^2 = 0.07$]: female students in comparison to male students more strongly perceived IT as allowing them to pursue agentic goals.

In the next step we compared IT and non-IT female professionals in terms of goal affordance (Hypothesis 2c and 2d). Analysis shows no significant difference between female employees in IT and in non-IT regarding communal goal affordance [$F_{(1,217)} = 2.77$, $p = 0.097$, $\eta^2 = 0.01$]. A different pattern was observed for students: female students in IT have lower communal goal affordance than non-IT students [$F_{(1,128)} = 5.24$, $p = 0.024$, $\eta^2 = 0.04$].

We also compared IT and non-IT male professionals. For communal goal affordance, the difference for employees was not significant [$F_{(1,206)} = 0.99$, $p = 0.320$, $\eta^2 = 0.01$], but it was for students [$F_{(1,102)} = 9.54$, $p = 0.003$, $\eta^2 = 0.09$]. As in the case of female, male students in IT have lower communal goal affordance than non-IT students.

For agentic goal affordance, the difference was significant for female employees [$F_{(1,217)} = 5.46$, $p = 0.020$, $\eta^2 = 0.03$]and students [$F_{(1,128)} = 8.26$, $p = 0.005$, $\eta^2 = 0.06$]: both IT female employees and students have higher agentic goals affordance than non-IT employees and students in the comparison group. For male employees, the difference in agentic goal affordance was not significant [$F_{(1,206)} = 0.08$, $p = 0.774$, $\eta^2 = 0.01$]. The same lack of effect was for male students [$F_{(1,102)} = 0.45$, $p = 0.505$, $\eta^2 = 0.01$].

Lastly, we compared female IT students and employees. Female students in IT have higher agentic goal affordance than employees [$F_{(1,153)} = 48.32$, p <0.001, $\eta^2 = 0.24$]. The opposite pattern was for communal affordance: female IT employees have a stronger belief that the domain allows them to pursue communal goals [$F_{(1,153)} = 10.30$, $p = 0.002$, $\eta^2 = 0.06$] than do the female students. Male students in IT have higher agentic goal affordance than do the male employees [$F_{(1,175)} = 17.49$, p <0.001, $\eta^2 = 0.09$], but this effect was not observed for communal congruency [$F_{(1,175)} = 2.79$, $p = 0.097$, $\eta^2 = 0.02$].

Sense of social belonging

First, we compared female and male employees in IT (Hypothesis 3a). The effect was significant [$F_{(1,203)} = 7.16$, $p = 0.008$, $\eta^2 = 0.03$]: contrary to our hypothesis, women in IT had higher sense of belonging to the field than men, but this effect was not significant for students [$F_{(1,125)} = 0.87$, $p = 0.352$, $\eta^2 = 0.01$].

TABLE 1 Results of a series of MANOVAs showing differences between male vs. female employees and male vs. female students in the IT field.

	IT employees							IT students						
	Female		Male					Female		Male				
	M	*SD*	*M*	*SD*	$F_{(1,203)}$	*p*	η^2	*M*	*SD*	*M*	*SD*	$F_{(1,127)}$	*p*	η^2
1. Communal endorsement	5.04	1.04	4.64	1.15	6.48*	0.012	0.03	4.36	1.07	3.99	1.09	3.66	0.058	0.03
2. Agentic endorsement	5.14	0.71	5.13	0.75	0.01	0.905	0.01	5.12	0.69	5.09	0.79	0.06	0.799	0.01
3. Communal affordance	4.61	1.07	4.34	1.18	2.92	0.089	0.01	4.10	0.92	4.03	1.03	0.22	0.693	0.01
4. Agentic affordance	4.84	0.97	4.67	1.03	1.48	0.225	0.01	5.78	0.65	5.35	0.93	9.15*	0.003	0.07
5. Social belonging	5.87	0.96	5.44	1.22	7.16*	0.008	0.03	5.08	1.45	5.31	1.28	0.87	0.352	0.01
6. Self-efficacy	5.30	0.84	5.22	0.91	0.33	0.566	0.01	4.26	0.87	4.61	0.95	4.77*	0.031	0.04

*p < 0.05. This table corresponds with hypotheses H 1a, H 1c, H 2a, H 2b, H 3a, H 4a.

TABLE 2 Results of a series of MANOVAs showing differences between female employees and students in IT vs. non-IT fields.

Dependent variable	Female employees							Female students						
	IT field		Non-IT field					IT field		Non-IT field				
	M	*SD*	*M*	*SD*	$F_{(1,217)}$	*p*	η^2	*M*	*SD*	*M*	*SD*	$F_{(1,128)}$	*p*	η^2
1. Communal endorsement	5.04	1.04	5.06	0.99	0.02	0.880	0.01	4.36	1.07	4.97	1.21	9.14*	0.003	0.07
2. Agentic endorsement	5.14	0.70	5.27	0.78	1.52	0.219	0.01	5.12	0.69	5.38	0.92	3.46	0.065	0.03
3. Communal affordance	4.61	1.07	4.34	1.22	2.77	0.097	0.01	4.10	0.92	4.51	1.16	5.24*	0.024	0.04
4. Agentic affordance	4.84	0.97	4.52	1.02	5.46*	0.020	0.03	5.78	0.65	5.33	1.11	8.26*	0.005	0.06
5. Social belonging	5.87	0.96	5.59	1.23	2.95	0.087	0.01	5.08	1.45	5.44	1.46	1.96	0.164	0.01
6. Self-efficacy	5.30	0.84	5.29	0.92	0.01	0.993	0.01	4.26	0.87	4.50	1.16	1.80	0.182	0.01

*p < 0.05. This table corresponds with hypotheses H 1b, H 2c, H 2d. H 3b, H 4b.

TABLE 3 Results of a series of MANOVAs showing differences between employees and students in the IT field.

Dependent variable	IT female							IT male						
	Employees		Students					Employees		Students				
	M	*SD*	*M*	*SD*	$F_{(1,153)}$	*p*	η^2	*M*	*SD*	*M*	*SD*	$F_{(1,175)}$	*p*	η^2
1. Communal endorsement	5.04	1.04	4.36	1.07	16.13**	<0.001	0.10	4.64	1.15	3.99	1.09	12.59**	<0.001	0.07
2. Agentic endorsement	5.14	0.70	5.11	0.69	0.03	0.858	<0.01	5.13	0.75	5.09	0.79	0.11	0.741	0.01
3. Communal affordance	4.61	1.07	4.10	0.92	10.30*	0.002	0.06	4.34	1.18	4.03	1.03	2.79	0.097	0.02
4. Agentic affordance	4.84	0.97	5.78	0.65	48.32**	<0.001	0.24	4.67	1.03	5.35	0.93	17.49**	<0.001	0.09
5. Social belonging	5.87	0.96	5.08	1.45	16.16**	<0.001	0.10	5.44	1.22	5.31	1.28	0.39	0.531	0.01
6. Self-efficacy	5.30	0.84	4.26	0.87	56.22**	<0.001	0.27	5.22	0.91	4.61	0.94	16.41**	<0.001	0.09

*p < 0.05 and **p < 0.001. This table corresponds with exploratory analysis.

The comparison of females in IT to those in the control group shows no difference for employees [$F_{(1,217)} = 2.95$, $p = 0.087$, $\eta^2 = 0.01$] and students [$F_{(1,128)} = 1.96$, $p = 0.164$, $\eta^2 = 0.01$]. There was also no significant effect for male employees [$F_{(1,206)} = 0.08$, $p = 0.773$, $\eta^2 = 0.01$] and male students [$F_{(1,102)} = 3.65$, $p = 0.059$, $\eta^2 = 0.04$].

Female employees in IT have a higher sense of belonging to the field than do the students [$F_{(1,153)} = 16.16$, $p < 0.000$, $\eta^2 = 0.10$). There was no such effect for men in IT [$F_{(1,175)} = 0.39$, $p = 0.531$, $\eta^2 = 0.01$].

Self-efficacy

Contrary to our hypothesis 4a, female employees in IT had the same level of self-efficacy as male employees [$F_{(1,203)} = 0.33$, $p = 0.566$, $\eta^2 = 0.01$]. What is interesting, this effect was not true for students [$F_{(1,125)} = 4.77$, $p < 0.031$, $\eta^2 = 0.04$]. Confirming our hypothesis, female students in IT had lower levels of self-efficacy than male students.

Moreover, women in IT had the same level of self-efficacy as women in the non-IT field, as was true for employees [F (1,217) $= 0.01$, $p = 0.993$, $\eta^2 < 0.01$] and students [F (1,128) $= 1.80$, $p = 0.182$, $\eta^2 = 0.01$]. There was no effect for male employees [$F_{(1,206)} = 0.56$, $p = 0.455$, $\eta^2 < 0.01$] and male students [F (1,102) $= 0.01$, $p = 0.966$, $\eta^2 < 0.01$].

What is interesting, female employees in IT had significantly higher self-efficacy than IT female students [$F_{(1,153)} = 56.22$, $p < 0.001$, $\eta^2 = 0.27$]. This effect was also observed for men in IT [$F_{(1,175)} = 16.41$, $p < 0.001$, $\eta^2 = 0.09$].

Data presented in this article are available here https://osf.io/pfq8s/?view_only=10c957f6632d42f49e83c38bee0d94e7.

General discussion

The European Commission and other world organizations are raising the alarm about the shortage of staff in the IT sector (European Commission, 2008; United Nations Technology and Innovation Labs Report, 2019; Eurofound, 2021). The inequality in terms of staff's gender is also a serious concern in high tech innovative companies since considerable amount of data show that adequate female representation in the workplace contributes to better outcomes of the company (Hunt et al., 2015).

In our study we sampled men and women in the IT sector on different levels of their professional life (students

TABLE 4 Means, standard deviations and correlations between main variables.

Variable	*M*	*SD*	1	2	3	4	5
1. Agentic endorsement	5.20	0.81	–				
2. Communality endorsement	4.72	1.11	0.124**	–			
3. Agentic affordance	4.93	1.08	0.233**	0.050	–		
4. Communality affordance	4.33	1.15	0.195**	0.372**	0.415**	–	
5. Self-efficacy	4.97	1.00	0.236**	0.207**	0.113**	0.265**	–
6. Belonging	5.49	1.28	0.080*	0.118**	0.209**	0.249**	0.489**

*p < 0.05, **p < 0.001.

TABLE 5 Means and confidence intervals for all dependent variables for female and male employees and students in IT and business fields.

	IT employees		Business employees		IT students		Business students	
	M_{female} [95% CI]	M_{male} [95% CI]	M_{female} [95% CI]	M_{male} [95% CI]	M_{female} [95% CI]	M_{male} [95% CI]	M_{female} [95% CI]	M_{male} [95% CI]
1. Communal endorsement	5.05ad	4.64bd	5.06a	4.71bd	4.36bc	3.99c	4.97d	4.52d
	[4.81, 5.27]	[4.46, 4.84]	[4.09, 5.24]	[4.48, 4.93]	[4.11, 4.61]	[3.70, 4.28]	[4.68, 5.25]	[4.22, 4.82]
2. Agentic endorsement	5.14ab	5.13ab	5.27ab	5.12a	5.11ab	5.09ab	5.38ab	5.48ab
	[4.98, 5.30]	[5.00, 5.26]	[5.12, 5.49]	[4.93, 5.13]	[4.93, 5.30]	[4.88, 5.30]	[5.17, 5.59]	[5.26, 5.69]
3. Communal affordance	4.61a	4.34ac	4.34abc	4.16bc	4.10c	4.03c	4.51d	4.62d
	[4.35, 4.88]	[4.12, 4.55]	[4.14, 4.55]	[3.90, 4.42]	[3.87, 4.33]	[3.76, 4.30]	[4.25, 4.78]	[4.34, 4.90]
4. Agentic affordance	4.84a	4.67a	4.52a	4.62a	5.78b	5.35c	5.33c	5.24c
	[4.62, 5.06]	[4.90, 4.85]	[4.33, 4.70]	[4.39, 4.86]	[5.56, 5.98]	[5.12, 5.59]	[5.10, 5.56]	[4.99, 5.48]
5. Social belonging	5.87a	5.44b	5.59ab	5.39b	5.08b	5.31b	5.44b	5.79a
	[5.61, 6.13]	[5.22, 5.64]	[5.39, 6.13]	[5.13, 5.64]	[4.76, 5.40]	[4.94, 5.40]	[5.08, 5.80]	[5.41,6.17]
6. Self-efficacy	5.30a	5.22ac	5.29a	5.13a	4.26b	4.61c	4.50bc	4.61bc
	[5.10, 5.49]	[5.07, 5.38]	[5.14, 5.45]	[4.93, 5.33]	[4.05, 4.47]	[4.37, 4.86]	[4.23, 4.77]	[4.32, 4.90]

Values in square brackets indicate the 95% confidence interval for each mean, lower-limit and upper-limit, respectively. The small letters indicate the differences between means.

and employees). Our goal was to study the possible gender differences in various aspects of the perception of the IT domain to inform potential interventions aiming at retaining women, who are already educated in this domain, in the IT sector. We expected to find specific gender differences and differences between IT and non-IT individuals. We focused on their goal endorsement and affordance, sense of social belonging, and self-efficacy. We were also interested in exploring the differences between students and women and men already employed in a given domain.

Women and men within IT and outside of IT

To compare women and men within IT and outside of IT, we examined the goal endorsement and affordance and tested for potential gender and group differences. We predicted that women in IT will value communal goals more than men in IT (H 1a), but less so than women in the non-IT sector (H 1b). Only the first assumption turned out to be true. Women in our IT employee sample indeed valued communal goals more than men yet they valued them equally as women in the non-IT group. Although it is important to note that this gender difference was not so prominent, with the men valuing communal goals only marginally less than women. This result is consistent with other findings showing that for women communal goals play an important role (Pyrkosz-Pacyna et al., 2019) and may in turn predict their engagement in specific lines of work, for example in the STEM fields (Cheryan et al., 2009). Our results also show that the opportunity to realize one's communal goals might be important for men in IT as well. However, we also did not show that women in IT care less about communal goals than women in other professions (H 1c). Therefore, perceiving IT as a domain where achieving communal goals is possible might be one of the factors contributing to greater female representation in IT but it can also appeal to men, as they, while pursuing work-life balance, might be also finding communal goals as appealing (Diekman et al., 2010; Croft et al., 2015; Steinberg and Diekman, 2018). We also wanted to see whether communal goals are more important than agentic goals for women in IT (H 1d). On the one hand, women in general tend to value communal goals slightly more than agentic goals (Wood and Eagly, 2002). On the other hand, perhaps women who choose IT do so because of their particular focus on agentic goals. We predicted that the first explanation is the more plausible one, yet our findings showed that women in IT valued agentic goals no less than they value communal goals. People stereotypically tend to think that women care less about agentic goals and simply strive for career development less than they care about reaching communal goals. Our results align with the increasing bulk of literature showing that agentic goals are becoming equally important for both women and men (Moore et al., 2008; Pyrkosz-Pacyna et al., 2019). We may thus suspect that women might be actually attracted to IT by perceiving it as allowing their agentic goals to be fulfilled - our study participants in both IT and in the non-IT group, did value agentic goals as well as the communal ones. Recruitment or promotion strategies omitting this aspect of women's values can influence future decisions regarding their engagement in STEM.

In the next step we wanted to see whether women in IT perceive their field as enabling them to achieve their goals. We hypothesized that women in IT will perceive less communal goal affordance than do the men in IT (H 2a) and the women in the control group (H 2c). Neither of our hypotheses were confirmed, showing that there might be no gender or interdomain difference between women in the IT and non-IT sectors in terms of communal goal affordance. It is worth mentioning here that in general the level of perceived communal and agentic affordance in the studied sample was at the medium level. When it comes to agentic goal affordance, again our hypotheses were not confirmed - men and women did not differ in perception of agentic goal affordance in IT (H 2b). This is a positive finding in the light of the fact that lack of agentic goal affordance was in previous studies found to be contributing to women dropping out from STEM domains (Diekman et al., 2010). When aiming for interdomain comparison, women in IT described their domain as affording agentic goals to a higher extent than did the women in the non-IT group (H 2d). This result is in accordance with our hypotheses. Even though it might seem women in male dominated fields might be less able to achieve their agentic goals, in our sample this was not the case. This finding is also of great practical significance - since agentic goals are increasingly important to women, the perception of IT as affording this striving might increase the likelihood of choosing an IT career path.

As for the sense of social belonging we predicted that women in IT will have a lower sense of social belonging than do the men in IT (H 3a) and the women in the control group (H 3b). Again, our assumptions were not entirely supported. Surprisingly, and contrary to previous findings, we found that women have a higher sense of belonging to the field than do the men in IT. This result may indicate that the IT domain might have become more welcoming toward women especially in the light of the many interventions directed at women to encourage them to study IT. Several studies (Cheryan and Plaut, 2010; Good et al., 2012; Tellhed et al., 2017; Aelenei et al., 2020) did show that indeed women tend to feel that they do not naturally belong in STEM but if their sense of fit was experimentally increased then they declared higher interest in career in STEM. Such campaigns are recognized at technical universities across Poland (see: IT for SHE or Girls as Engineers! & Girls go Science! Campaigns) and they may have increased the visibility of women pursuing IT careers.

Efforts toward validating the effectiveness of interventions might shed a light on the changes that need to be implemented to increase their efficiency.

Finally, we focused on women's vs men's assumed lower self-efficacy. Yet again in our study we did not find any significant gender or interdomain differences, thus proving that women in IT have a self-efficacy similar to that of men in IT (H 4a) and as that of women in the comparison group (H 4b).

Taken together, as shown by these results, we were unable to show many of the anticipated gender differences that might prove that there are in fact fewer gender-based differences than is assumed. Men and women in IT seem to be alike in terms of their perception of goal affordance, sense of belonging, and self-efficacy. In fact, in the IT sample, the women even outscored men in terms of the sense of social belonging to their workplace. All these findings, in our opinion, support a strong need for continuous monitoring of the gender (in)balance to be able to design and implement the interventions that address specific problems with appropriate, evidence-based solutions.

Employees and students within IT

Our second line of analysis focused on exploring similarities and differences among female IT employees and students. Our data showed that in the study sample female IT students valued communal goals to a lesser degree than do women already working in IT. The same effect was also visible for male IT workers and students, with male employees valuing communal goals significantly higher than do the IT students. When it comes to agentic goals, we found that they were equally important for all groups, namely IT male and female workers and students, and in all cases more so than communal goals. These results indicate that there are noticeable differences in goal endorsement on different career levels in terms of communal goals - they seem to be more important for employees than they do for students. This difference might be due either to generational factors or to a change in attitude over time, which would be possible to examine only with longitudinal research. We argue that this result is reflecting the current focus of students, both male and female, which is to achieve agentic work goals. Only when these goals are secured employees start to focus on communal goals, perhaps due in part to work conditions that require tapping into these competences, i.e. cooperation, teamwork or communication. Our results are congruent with some results present in the literature concerning IT employees indicating that agentic goals are becoming equally important for men as well as women (Moore et al., 2008; Pyrkosz-Pacyna et al., 2019) and thus recruitment or promotion strategies failing to underline the agentic aspect of women's values, along with communal ones, might not be fully effective in attracting female candidates.

We also found a difference between students and employees in terms of their perception of goal affordance. Women already working in IT found that it better afforded the attainment of communal goals than did the IT students. The opposite was true for agentic goals affordance. These results show that again, there is a significant difference in perception of the IT domain by those already familiar with it and those yet to be employed there.

Taken together, female IT students value agentic goals more than they value communal goals (contrary to common gender stereotypes) and at the same time perceive greater possibility to achieve those goals in their future workplace than do the men in IT and the women in the comparison group. Although research within STEM shows that STEM occupations are perceived as affording communal goals to a lesser extent that other professions (see e.g., Diekman et al., 2010) the IT sector may signalize that communal goal congruence is achievable among female employees (Herz, 1997; Sanger et al., 1997; Clayton et al., 2009; Santos et al., 2022).

These results should indicate a higher probability of women to engage in IT after graduation. They also suggest that information on goal attainment in a given workplace might be a relevant practical strategy to communicate with potential candidates. At the same time, monitoring the goal strivings and goal attainment among men and women already working in IT might provide a significant advantage when it comes to achieving gender balance in this sector.

When it comes to social belonging, female IT students revealed levels similar to those of their male colleagues and the female students in the comparison group. It would appear that once women get into IT, they tend to enjoy a similar level of sense of social belonging than do the other study participants. We did however find a significant gender difference when it comes to IT student's self-efficacy, namely that women have considerably lower self-efficacy than men. Lower self-efficacy than that among men might be a crucial ingredient when searching for the origins of female underrepresentation in IT. Many studies thus far have shown how low self-efficacy may influence decisions regarding future career goals in the context of women in STEM (Correll, 2001; Singh et al., 2007; Good et al., 2012; Sterling et al., 2020; Stewart et al., 2020). It is important to note that we have found no difference in self-efficacy among female students in IT and those in the comparison group. We suspect that IT students might reveal lower self-efficacy, as they are subjected to constant comparison with male counterparts and various forms of gender stereotypes. However, this turned out to be a false assumption. Also, it is important to note, that the self-efficacy gender difference was not present among the employees sample. This indicates that the highly vulnerable period for women when it comes to building self-efficacy is during the process of gaining education. This result highlights that more attention to building self-efficacy among IT students and longitudinal research would be recommended to allow for more precise insights into factors contributing to lower female representation in STEM over time.

Based on our findings and previous literature we prepared a list of practical implications to be taken into consideration for policymakers and individuals involved in the process of designing and delivering gender diversity in STEM interventions.

Practical implications for policymakers and interventions designers

Tailoring problem-focused interventions for different stages of education and career

Our study showed that there is a significant difference between male and female IT students in terms of self-efficacy, namely the self-efficacy among female IT students is lower than among male students. This effect was not however found among IT employees suggesting that different interventions are due for different moments in one's career trajectory. For students more attention to building their self-efficacy would be needed whereas less so for employees.

Importance of monitoring of longitudinal effects

Our research showed that goal endorsement among students and employees is different. For example, we discovered that communal and agentic goal endorsement is equally important for IT female employees but not for IT female students—for IT female students agentic goals are significantly more important. Again, these results call for different intervention strategies but also, they point out to the importance of conducting longitudinal research focused on women in STEM fields. Longitudinal research enables more precise insights into factors contributing to lower female representation in STEM over time.

Taking into account socio-economic and geographical context of intervention beneficiaries

Most studies regarding women in STEM are conducted in western cultures. There is also considerable literature regarding women in STEM professions in developing economies (e.g., Stoet and Geary, 2018; Huang et al., 2020). These studies point out to some important differences among studied mechanisms for example Stoet and Geary (2018) using international database of over half a million adolescents' achievement in science, mathematics and reading showed that sex differences in the academic readiness to pursuit STEM careers favoring boys increase with gender equality levels. It is therefore important to foster more research including various country-level variables and economic contexts to find and test solutions suitable for specific contexts.

Providing women with insights into the specifics of pursuing an IT career

As our research showed, the perception of work in IT is in some respects significantly different among female students and female employees. Therefore, interventions aimed at showcasing careers in various STEM disciplines, including IT, would be beneficial for acquiring an accurate and informed view on the matter. Such interventions may include among others: guest lectures, open days, mentoring programs, video testimonials, internships, and other. Especially mentoring has been proven to be an effective tool in supporting women in STEM (Dennehy and Dasgupta, 2017).

Investigating and pilot testing various interventions aimed at boosting self-efficacy

Our research, as well as numerous others, showed that women tend to have lower levels of self-efficacy than men. The influence of low self-efficacy on educational and vocational decisions is well documented (e.g., Dennehy and Dasgupta, 2017; Easterbrook et al., 2021). Specifically in highly male-dominated fields women's low self-efficacy is of vast importance. Therefore, it is crucial to design and implement interventions aimed at raising the level of self-efficacy among women, especially students since as our study showed, the gender difference seems to not appear among working individuals as they do among students.

Continuous monitoring of gender related mechanisms in STEM fields

Much valuable research to this date highlighted the importance of social belonging when it comes to female representation in STEM (e.g., Cheryan and Plaut, 2010; Diekman et al., 2010). These results emphasize the value of both theory and evidence-based intervention aimed at attracting and keeping women in STEM. Since these interventions are strongly contextualized, for example by time and target group, even when proving effective in one context might not be as effective in others (Easterbrook et al., 2021). Hence, they need to be tailored by focusing on the needs of the given target group in a given time and based on proper diagnosis of the needs of the group at which the intervention is aimed. With development of gender equality plans in different education contexts currently happening across academic institutions across the world new measures and new goals might have to be established to improve gender balance in different fields. For example, in our study we hypothesized that women in the IT occupation will have a lower sense of social belonging than men in similar positions. This assumption turned out not to be true. Quite the opposite - surprisingly, in our sample women in IT positions had a higher sense of social belonging than men. Results that we gathered suggest that at least in some cases this effect might be currently less pervasive. In our opinion continuous replication of such culturally sensitive mechanisms would be beneficial for achieving more accurate and timely data.

Limitations and future research

In general, our findings highlight two important factors relating to women's presence in the IT sector. First, one cannot generalize the results obtained among students to the employee population. We found considerable differences in almost all the measured aspects when comparing students and employees' samples within IT. Second, the results indicate that there might be a potential shift in perception of IT employment when moving up the occupational ladder. Future research in this area might be very useful in the effort to establish whether there is an actual change of goals and beliefs. A longitudinal study paradigm would be especially beneficial for further analysis especially in terms of investigating if the discovered differences between researched groups stem from changes in experience or are generational differences or even a result of comparing two distinct age groups.

Our study is not free from limitations. Firstly, the participants of our study were recruited from two IT companies, so our findings might simply reflect the given companies' focus on the inclusion of women. Indeed, both companies avow such policies. Furthermore, the choice of banking and management as comparative fields does not allow for a full comparison with the non-STEM field. Secondly, even though the underrepresentation of women in IT is rather universal country-wise, we do observe some country-level variations. A recent study focusing on 41 countries in the OECD and EU shows that the percentage of women working in the information and communication technology sector varies from 9 to 11% in countries with lowest representation of women in IT (Slovak Republik, Turkey and Israel) up to maximum 24-30% in countries with the highest representation of women in IT (Bulgaria, Australia, and Romania; Honeypot, 2018). In a study by Stoet and Geary (2018), the authors point out the paradox of less developed countries having a surprisingly high number of women in STEM, presumably due to the perception of STEM as providing good socio-economic improvement possibilities. As in many other cases, the vast preponderance of research related to lower representation of women in STEM is conducted in the US and other developed countries. As this has been conducted in Poland, a country with relatively little experience with inclusion programs within IT companies and very traditional when it comes to gender stereotypes, our study contributes to the understanding of IT representation within a broader cultural sample. In terms of methodological improvements in future research measures, specifically addressing work in IT (in our study we asked for participants' line of work) might be more accurate.

Conclusions

Our findings contribute to current research lines focused on lower representation in the IT sector by adding another under-researched context of Poland (most of the current studies are conducted in the US or in other western countries). Although there is an abundance of job offers and career possibilities in IT for both men and women, companies struggle to attract women to IT positions. The practical implication of our study is the broadened understanding of the factors contributing to the low representation of women in IT, specifically in terms of the goal congruence perception with its impact on the sense of belonging and thus on the sense of engagement with work. Our results show that women might feel welcomed and included in the IT sector.

Gender diversity translates to a better fit between technology and society. That women have almost no voice in the creation of some important technological innovations is detrimental to the industry and society (Selby et al., 1997). Yet, 88 % of information technology patents are invented by male-only teams. Lastly, the IT industry is one of the best paying in the economy (Hays, 2020). Consequently, low female representation in IT contributes to an elevated gender pay gap Fry et al. (2021). Studies show that the gender pay gap in STEM does exist although it is smaller than in other sectors: 14 % in STEM vs. 20 % in non-STEM fields (Beede et al., 2011; Jasko et al., 2020), which contributes to a narrowing of the pay gap ratio. Some researchers argue that the pay gap might be even larger in IT due to persisting stereotypes and various well-known mechanisms such as the sticky floor effect - discriminatory employment pattern keeping mainly women in the lower positions with low mobility and invisible barriers to their career advancement (Segovia-Pérez et al., 2020). And lastly, IT is perceived rather favorably by the IT workforce as a good place to work, also by women - in the study by Hewlett et al. (2008) 80 % of women report 'loving their work'. Our results also seem to confirm the assertion which should be disseminated more widely to attract more women within IT. They do like it. They do feel they belong there.

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 in the article/supplementary material.

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 required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

KD organized the database and performed the statistical analysis. JP-P wrote the first draft of the manuscript. NK-B provided essential corrections to the sections of the manuscript. All authors contributed to the conception and design of the study, manuscript revision, read, and approved the submitted version.

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

Aelenei, C., Martinot, D., Sicard, A., and Darnon, C. (2020). When an academic culture based on self-enhancement values undermines female students' sense of belonging, self-efficacy, and academic choices. *J. Soc. Psychol.* 160, 373–389. doi: 10.1080/00224545.2019.1675576

Ashcraft, C., McLain, B., and Eger, E. (2016). *Women in Tech: The Facts.* National Center for Women and Technology (NCWIT).

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. NY: W.H. Freeman.

Baumeister, R. F., and Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* 117, 497. doi: 10.1037/0033-2909.117.3.497

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., and Doms, M. E. (2011). Women in STEM: a gender gap to innovation. *Econ. Stat. Admin. Issue Brief* 4–11.

Brainard, S., and Carlin, L. (1998). A six-year longitudinal study of undergraduate women in engineering and science. *J. Eng. Educ.* 87, 369–375. doi: 10.1002/j.2168-9830.1998.tb00367.x

Brown, E. R., Thoman, D. B., Smith, J. L., and Diekman, A. B. (2015). Closing the communal gap: The importance of communal affordances in science career motivation. *J. Appl. Soc. Psychol.* 45, 662–673. doi: 10.1111/jasp.12327

Cheryan, S., and Plaut, V. C. (2010). Explaining underrepresentation: a theory of precluded interest. *Sex Roles*, 63, 475–488. doi: 10.1007/s11199-010-9835-x

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, 1045–1060. doi: 10.1037/a0016239

Clayton, K. L., Von Hellens, L. A., and Nielsen, S. H. (2009). "Gender stereotypes prevail in ICT: a research review," in *Proceedings of the Special Interest Group on Management Information System's 47th Annual Conference on Computer Personnel Research* (pp. 153–158). doi: 10.1145/1542130.1542160

Correll, S. J. (2001). Gender and the career choice process: the role of biased self-assessments. *Am. J. Sociol.* 106, 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. doi: 10.1177/000312240406900106

Cowgill, C., Halper, L., Rios, K., and Crane, P. (2021). "Why So Few?": differential effects of framing the gender gap in STEM recruitment interventions. *Psychol. Women Q.*, 45, 61–78. doi: 10.1177/03616843209 65123

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: 10.1177/1088868314 564789

Dennehy, T. C., and Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proc. Nat. Acad. Sci. USA* 114, 5964–5969.

Diekman, A. B., Brown, E. R., Johnston, A. M., and Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychol. Sci.* 21, 1051–1057.

Dweck, C. S. (2007). "Is Math a Gift? Beliefs That Put Females at Risk," in *Why Aren't More Women in Science?: Top Researchers Debate the Evidence*, eds Ceci, S. J., and Williams, W. M. (American Psychological Association), 47–55. doi: 10.1037/11546-004

Easterbrook, M. J., Harris, P. R., and Sherman, D. K. (2021). Self-affirmation theory in educational contexts. *J. Social Issues* 77, 683–701. doi: 10.1111/j.1471-6402.1987.tb00781.x

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychol. Women Q.* 11, 135–172.

Eurofound (2021). *Tackling Labour Shortages in EU Member States.* Luxembourg: Publications Office of the European Union.

European Commission (2008). *Women in ICT: Status and the Way Ahead*. Brussels, DG Information Society and Media.

European Commission Directorate-General for Research and Innovation (2021). *She Figures 2021: Gender in Research and Innovation: Statistics and Indicators*. Publications Office. Available online at: https://data.europa.eu/doi/10.2777/06090

Fouad, N. A., Singh, R., Fitzpatrick, M. E., and Liu, J. P. (2012). *Stemming the Tide: Why Women Leave Engineering.* Milwaukee, WI: University of Wisconsin.

Fry, R., Kennedy, B., and Funk, C. (2021). *STEM Jobs See Uneven Progress in Increasing Gender, Racial and Ethnic Diversity*. Washington, DC: Pew Research Center.

Good, C., Rattan, A., and Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *J. Personal. Soc. Psychol.* 102, 700–717. doi: 10.1037/a0026659

Gu, J. (2018). *Women Lose Out to Men Even Before They Graduate From College.* Available online at: https://www.bloomberg.com/graphics/2018-women-professional-inequality-college/

Hays (2020). *Annual Report and Financial Statements.* Available online at: http://fnope.org.pl/dokumenty/2020/01/Hays_Raport_placowy_2020.pdf

Herz, J. C. (1997). *Joystick Nation*. London: Abacus.

Hewlett, S. A., Luce, C. B., Servon, L. J., Sherbin, L., Shiller, P., Sosnovich, E., and Sumberg, K. (2008). The Athena factor: reversing the brain drain in science, engineering, and technology. *Harv. Bus. Rev. Res. Rep.* 10094, 1–100.

Honeypot (2018). *2018 Women in Tech Index*. Available online at: https://honeypotio.github.io/women-in-tech/ (accessed June 1, 2022).

Huang, J., Gates, A. J., Sinatra, R., and Barabási, A. L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proc. Nat. Acad. Sci. USA* 117, 4609–4616. doi: 10.1073/pnas.1914221117

Hunt, V., Layton, D., and Prince, S. (2015). *Diversity Matters*. McKinseyandCompany. Available online at: https://www.mckinsey.com/~/media/mckinsey/business%20functions/people%20and%20organizational%20performance/our%20insights/why%20diversity%20matters/diversity%20matters.pdf (accessed June 1, 2022).

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. Social Issues* 76, 512–542. doi: 10.1111/josi.12392

Knapińska, A. (2022). *Women in Politechnics Report. Państwowy Instytut Badawczy.* Available online at: http://www.dziewczynynapolitechniki.pl/pdfy/raport-kobiety-na-politechnikach-2022.pdf (accessed June 1, 2022).

Kossowska, M., Górniak, J., Prokopowicz, P., Zmuda, G., Malinowska, D., Pyrkosz, J., et al. (2012). *Study of Competences in BPO and ITO in Kraków. Final Report on the Conducted Research*. Available online at: https://www.krakow.pl/zalacznik/21521 (accessed June 1, 2022).

Kost-Smith, L. E., Pollock, S. J., and Finkelstein, N. D. (2009). Characterizing the gender gap in introductory physics. *Phys. Rev. Special Top. Phys. Educ. Res.* 5, 1–14. doi: 10.1103/PhysRevSTPER.5.010101

Miura, I. T. (1987). The relationship of computer self-efficacy expectations to computer interest and course enrollment in college. *Sex Roles* 16, 303–311. doi: 10.1007/BF00289956

Miura, I. T. (2020). Computer self-efficacy: a factor in understanding gender differences in computer course enrollment. *IEEE Technol. Soc. Mag.* 3, 43–48. doi: 10.1109/MTAS.2003.1237471

Moore, K., Griffiths, M., Richardson, H., and Adam, A. (2008). Gendered futures? Women, the ICT workplace and stories of the future. *Gender Work Organ.* 15, 523–542. doi: 10.1111/j.1468-0432.2008.00416.x

Peña, M., Olmedo-Torre, N., Mas de Les Valls, E., and Lusa, A. (2021). Introducing and evaluating the effective inclusion of gender dimension in STEM higher education. *Sustainability* 13, 4994. doi: 10.3390/su130 94994

Pyrkosz-Pacyna, J., Dukała, K., Szastok, M. (2019). Men and Women on the STEM track. Similar or diverse? *Stud. Human.* 4, 31–44. doi: 10.7494/human.2019.18.4.31

Ramsey, L. R., Betz, D. E., and Sekaquaptewa, D. (2013). The effects of an academic environment intervention on science identification among women in STEM. *Soc. Psychol. Educ.* 16, 377–397. doi: 10.1007/s11218-013-9218-6

Sanger, J., Wilson, J., Davies, B., and Whitakker, R. (1997). *Young Children, Videos and Computer Games: Issues for Teachers and Parents.* London: Falmer Press.

Santos, E. D., Albahari, A., Díaz, S., and de Freitas, E. S. (2022). Science and technology as feminine: raising awareness about and reducing the gender gap in STEM careers. *J. Gender Stud.* 31:4, 505–518, doi: 10.1080/09589236.2021.1922272

Segovia-Pérez, M., Castro Núñez, R. B., Santero Sánchez, R., and Laguna Sánchez, P. (2020). Being a woman in an ICT job: an analysis of the gender pay gap and discrimination in Spain. *New Technol. Work Employ.* 35, 20–39. doi: 10.1111/ntwe.12145

Selby, L., Young, A., and Fisher, D. (1997). "Increasing the participation of women in tertiary level computing courses: what works and why," in *Proceedings* *of ASCILITE 97, 14th Annual Conference of the Australian Society for Computers in Tertiary Education. Perth*, 8–10 December.

Singh, K., Allen, K. R., Scheckler, R., and Darlington, L. (2007). Women in computer-related majors: a critical synthesis of research and theory from 1994 to 2005. *Rev. Educ. Res.* 77, 500–533. doi: 10.3102/0034654307309919

Singh, R., Fouad, N. A., Fitzpatrick, M. E., Liu, J. P., Cappaert, K. J., and Figuereido, C. (2013). Stemming the tide: predicting women engineers' intentions to leave. *J. Vocat. Behav.* 83, 281–294. doi: 10.1016/j.jvb.2013.05.007

Statista (2021). *Women's Representation in Big Tech*. Available online at: http://www.statista.com/chart/4467/female-employees-at-tech-companies/ (accessed June 1, 2022).

Steinberg, M., and Diekman, A. B. (2017). Elevating positivity toward STEM pathways through communal experience: the key role of beliefs that STEM affords other-oriented goals. *Anal. Soc. Issues Public Policy* 17, 235–261. doi: 10.1111/asap.12135

Steinberg, M., and Diekman, A. B. (2018). Considering "why" to engage in STEM activities elevates communal content of STEM affordances. *J. Exp. Soc. Psychol.* 75, 107–114. doi: 10.1016/j.jesp.2017.10.010

Sterling, A. D., Thompson, M. E., Wang, S., Kusimo, A., Gilmartin, S., and Sheppard, S. (2020). The confidence gap predicts the gender pay gap among STEM graduates. *Proc. Natl. Acad. Sci. USA.* 117, 30303–30308. doi: 10.1073/pnas.2010269117

Stewart, J., Henderson, R., Michaluk, L., Deshler, J., Fuller, E., and Rambo-Hernandez, K. (2020). Using the social cognitive theory framework to chart gender differences in the developmental trajectory of STEM self-efficacy in science and engineering students. *J. Sci. Educ. Technol.* 29, 10231334. doi: 10.1007/s10956-020-09853-5

Stoet, G., and Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychol. Sci.* 29, 581–593. doi: 10.1177/0956797617741719

Stout, J. G., Dasgupta, N., Hunsinger, M., and McManus, M. A. (2011). STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *J. Pers. Soc. Psychol.* 100, 255. doi: 10.1037/a0021385

Tellhed, U., Bäckström, M., and Björklund, F. (2017). Will I fit in and do well? The importance of social belongingness and self-efficacy for explaining gender differences in interest in STEM and HEED majors. *Sex Roles*, 77, 86–96. doi: 10.1007/s11199-016-0694-y

Terrell, J., Kofink, A., Middleton, J., Rainear, C., Murphy-Hill, E. R., and Parnin, C. (2016). Gender bias in open source: pull request acceptance of women versus men. *PeerJ Preprints* 4, e1733v2. doi: 10.7287/peerj.preprints.1733v2

U.S. Department of Education. (2007). *The Nation's Report Card: America's high school graduates: Results from the 2005 NAEP high school transcript study, by C. Shettle, S. Roey, J. Mordica, R. Perkins, C. Nord, J. Teodorovic, J. Brown, M. Lyons, C. Averett, and D. Kastberg*. (NCES 2007- 467). Washington, DC: Government Printing Office.

United Nations Technology and Innovation Labs Report (2019). *Inclusion and Diversity: Tech It or Leave IT*. Available online at: https://until.un.org/sites/default/files/inline-files/UNTIL_Inclusion_and_Diversity_Tech_Report_2019.pdf (accessed June 1, 2022).

Walton, G. M., and Cohen, G. L. (2007). A question of belonging: race, social fit, and achievement. *J. Pers. Soc. Psychol.* 92, 82. doi: 10.1037/0022-3514.92.1.82

Women in IT. (2022). *Chances and Opportunities for Starting and Developing a Career. Gdynia: No Fluff Jobs sp. z o.o*. Available online at: https://nofluffjobs.com/insights/report-women-in-it-2022-chances-and-opportunities-for-starting-and-developing-a-career/ (accessed June 1, 2022).

Wood, W., and Eagly, A. H. (2002). A cross-cultural analysis of the behavior of women and men: implications for the origins of sex differences. *Psychol. Bull.* 128, 699–727. doi: 10.1037/0033-2909.128.5.699

frontiers | Frontiers in Psychology

TYPE Systematic Review
PUBLISHED 11 October 2022
DOI 10.3389/fpsyg.2022.954996

Check for updates

OPEN ACCESS

EDITED BY
Cheng Yong Tan,
The University of Hong Kong, Hong Kong SAR, China

REVIEWED BY
Beatrice Avolio,
Pontifical Catholic University of Peru, Peru
Dario Maldonado,
Universidad de los Andes, Colombia
Jessica To,
Nanyang Technological University, Singapore

*CORRESPONDENCE
Milagros Sáinz
msainzi@uoc.edu

SPECIALTY SECTION
This article was submitted to Gender, Sex and Sexualities, a section of the journal Frontiers in Psychology

RECEIVED 27 May 2022
ACCEPTED 15 August 2022
PUBLISHED 11 October 2022

CITATION
Sáinz M, Fàbregues S, Romano MJ and López B-S (2022) Interventions to increase young people's interest in STEM. A scoping review.
Front. Psychol. 13:954996.
doi: 10.3389/fpsyg.2022.954996

Interventions to increase young people's interest in STEM. A scoping review

Milagros Sáinz[1]*, Sergi Fàbregues[2], María José Romano[1] and Beatriz-Soledad López[1]

[1]Internet Interdisciplinary Institute, Universitat Oberta de Catalunya, Barcelona, Spain, [2]Department of Psychology and Education, Universitat Oberta de Catalunya, Barcelona, Spain

The underrepresentation of young people and particularly young women in many STEM fields has inspired various intervention programmes and research intended to boost their interest in these areas. The purpose of this scoping review is to examine the characteristics and effectiveness of interventions designed to encourage interest in STEM among secondary school students, particularly female students, over the past 20 years. A systematic search of the literature in five databases and additional search strategies resulted in identifying 215 studies evaluating interventions in different disciplinary fields. Data extraction and synthesis of these studies were carried out, focusing on the methodologies and theoretical foundations used. Twenty-five exemplars were selected to illustrate best practices in designing and evaluating interventions that address the various facets of young people's lack of interest in STEM. These interventions attempt to modify and/or manipulate multiple environmental and school factors to impact students' personal factors associated with STEM interest, such as achievement, self-perception of ability, and self-efficacy. Implications for the design of future interventions and potential outcomes are then discussed.

KEYWORDS

interest, interventions, gender, secondary education, stem

Introduction

The under representation of women in some Science, Technology, Engineering, and Mathematics (STEM) disciplines, such as the physical sciences, computer science, and engineering, is common throughout Western countries (UNESCO, 2018; Sáinz, 2020). Despite numerous efforts to promote equal opportunities, several studies have confirmed the persistence of sexist beliefs regarding the competences that men and women must possess to access and develop certain academic and professional activities (Leaper and Brown, 2014; Sáinz et al., 2020). These beliefs, which discourage young women from pursuing non-traditional STEM career pathways, revolve around the idea that women do not have sufficient technological and mathematical capabilities (Wang and Degol, 2013; Sáinz et al., 2020).

Many initiatives and interventions (i.e., empirical investigations that manipulate an independent variable and follow the effect over time) have been conducted worldwide to engage young women in STEM, especially in those STEM disciplines with a higher under-representation of women (Liben and Coyle, 2014; Rosenzweig and Wigfield, 2016; van den Hurk et al., 2019; Prieto-Rodriguez et al., 2020). Research has also highlighted the importance of analyzing STEM disengagement in order to improve the design of interventions, thereby attracting and retaining women in STEM pathways (Rosenzweig and Wigfield, 2016).

Interventions tend to represent operationalized theory in action. For some authors, they characterize the testing of a theory as applied in a given educational context (Lazowski and Hulleman, 2016). From a practical point of view, intervention-based studies expand our understanding of which intervention components are most effective in raising students' interest in STEM and how this can be sustained in the long term (Lazowski and Hulleman, 2016; Rosenzweig and Wigfield, 2016). Such understanding can guide educational policies and provide recommendations informed by scientific evidence (Liben and Coyle, 2014; Lazowski and Hulleman, 2016; Rosenzweig and Wigfield, 2016). For this reason, this review builds on previous systematic reviews of STEM intervention studies. These previous reviews have emphasized the importance of drawing on clear theoretical frameworks when designing and implementing a proper evaluation of an intervention's effectiveness, and how the various intervention features and components take effect (Rosenzweig and Wigfield, 2016; van den Hurk et al., 2019; Prieto-Rodriguez et al., 2020).

There is no single factor that alone can influence on girls' and women's participation, achievement, and progression in STEM education (Wang and Degol, 2013; UNESCO, 2018). For this reason, van den Hurk et al. (2019) in their review have categorized STEM intervention studies according to the factors they address: environmental level (such as stereotypical cultural and societal beliefs about gender and STEM, or the lack of female role models in STEM); school level (such as educational policies, school climate, teachers' beliefs and attitudes, or pedagogy); and student level (including cognitive characteristics such as academic ability and achievement, background characteristics such as gender and socioeconomic status, or affective characteristics, such as self-efficacy, motivation, belonging, and engagement). All these factors not only correlate with interest and persistence in STEM education, but they are also interrelated (Blickenstaff, 2005).

Personal level factors involved in shaping young people's (particularly girls') engagement and interest in STEM have been prioritized and measured by many intervention studies analyzed in prior reviews (Rosenzweig and Wigfield, 2016; van den Hurk et al., 2019; Prieto-Rodriguez et al., 2020). In this regard, Rosenzweig and Wigfield (2016) review of 53 intervention studies published between 1985 and 2015 focused on the following five motivation-related categories: competence-related beliefs (such as self-concept of ability, self-efficacy, confidence, and outcome expectations); beliefs about value, interest, and intrinsic motivation; attributions about academic success and failure; beliefs about intelligence; and achievement goal orientation. Many intervention studies address personal factors related to students' performance and engagement in STEM through changes either in school level factors, social-environmental level factors, or a combination of them. Interestingly, Prieto-Rodriguez et al. (2020) systematic review concluded that successful activities encouraging girls' STEM identity formation combined both inclusive curriculum and pedagogies (strategies at the school level) and exposure to female role models (strategy at the environment level).

Current societal stereotypes about the type of students who are expected to succeed in STEM (e.g., middle-class white male students) discourage many students who do not meet these attributes (e.g., girls, students from low SES or migrant families, as well as non-white students or students with disabilities) from entering in STEM fields (Good et al., 2003; Rosenzweig and Wigfield, 2016; Sáinz and Müller, 2018). However, in Western societies and contrary to students from low-SES families or with migrant and ethnical backgrounds, girls are not a minority in the school context (UNESCO, 2021). Moreover, females in most Western countries are highly represented in STEM disciplines that align with the caring role associated with feminine roles, such as medicine, chemistry, or biology (UNESCO, 2018).

In addition, several meta-reviews and reviews (i.e., Wang and Degol, 2013; van den Hurk et al., 2019) suggest the influence of school level factors in the teaching and learning of STEM subjects on girls' engagement, achievement, and progression in STEM. The instructional approach of STEM teachers, the curriculum of STEM subjects, or teachers' beliefs, attitudes, behaviors, and interactions with students can positively influence girls' performance and engagement with STEM education and their interest in pursuing STEM careers (UNESCO, 2018). Similarly, female students are frequently attributed less competence by their male peers in STEM activities developed in the classroom (Sáinz, 2020).

While some of these prior reviews have laid the theoretical groundwork for the design of present and future intervention studies (Rosenzweig and Wigfield, 2016; van den Hurk et al., 2019), they have failed to provide a comprehensive account of the alignment between the different methods and theories used to raise young people's interest in STEM. Moreover, previous reviews lack an in-depth analysis of the range of methods and methodological approaches used to evaluate the effectiveness of interventions. In the present review, we attempt to fill this literature gaps by focusing on the methods and theories used in intervention studies to increase young people's interest in STEM. In addition, this review will focus on intervention studies aimed at increasing girls' STEM motivation, while it also considers the intersection of gender with other inequality variables (such as race/ethnicity, and SES level). The findings of this review will shed light on how to enhance the design and implementation of interventions aimed at closing the gender gap in STEM pathways.

Challenges to the evaluation of the interventions

Several systematic reviews of interventions have highlighted a need for research into the effectiveness of intervention programmes intended to attract and retain highly motivated students in STEM fields (van den Hurk et al., 2019; Kolne and Lindsay, 2020; Prieto-Rodriguez et al., 2020). However, van den Hurk's (2019) systematic review of empirical studies on the effectiveness of STEM-related interventions published between 2005 and 2017, raised an important issue: only a few of these evaluations were adequately designed to determine whether the observed effects were actually caused by the intervention (van den Hurk et al., 2019). In many of the instances, the studies under review were neither randomly selected nor applied a control group.

Liben and Coyle's (2014) review of gender developmental interventions addressing the STEM gender gap provided a taxonomy of five intervention goals (remediate, revise, refocus, re-categorize, and resist) designed to enhance the alignment between (a) cognitive, personal, and/or perceived qualities of girls and women and (b) the demands and opportunities of STEM. In agreement with this review, many of the identified interventions were not systematically evaluated and therefore provided little empirical evidence of whether they successfully engaged girls and women in STEM-related subjects, especially in the long term. In this sense, it has been acknowledged that long-term interventions can contribute not only to raising, but also to maintaining young people's interest in STEM (Liben and Coyle, 2014; van den Hurk et al., 2019). According to Harackiewicz and Priniski (2018), the primary outcomes targeted by an intervention may not only serve as a measure of efficacy but can also trigger positive recursive processes that drive longer-term impacts. In fact, all the systematic reviews in the literature concluded that long-term interventions or repeated participation in interventions were most likely to result in meaningful engagement in STEM (Liben and Coyle, 2014; Rosenzweig and Wigfield, 2016; van den Hurk et al., 2019; Prieto-Rodriguez et al., 2020).

In their review of psychosocial-based interventions in higher STEM education, Harackiewicz and Priniski (2018) identified the following psychological processes as critical for various educational outcomes in higher education: students' lack of interest in certain STEM topics and subjects; students' lack of confidence in their own abilities; students experiencing identity threat in certain fields; students doubting about the suitability of an academic discipline, or about the fact that they belong to a particular STEM career pathway; students experiencing a cultural mismatch between institutional norms and their own values; and students suffering from various emotional issues.

In addition, Rosenzweig and Wigfield (2016) review on STEM motivation interventions also discussed about the need of understanding the impact of individual and contextual factors (moderators) to better disentangle their influence on the effects of interventions. In a similar fashion, Kolne and Lindsay (2020) systematic review analyzed the impact of programmes and interventions in increasing interest and participation in STEM education and careers among children and young people with disabilities. These authors concluded that more controlled designs are needed to determine the impact of specific intervention components and participant characteristics, such as gender and students' disabilities, on the evaluation of the intervention effectiveness.

The present review

The purpose of this scoping review is to examine the characteristics and content of intervention studies aimed at increasing young people's participation in STEM (female students in particular), conducted in various geographical areas over the past 20 years. The present review builds on prior systematic reviews of STEM interventions that have identified strategies for change that emerge from outstanding theories (Liben and Coyle, 2014; Rosenzweig and Wigfield, 2016). Additionally, it expands on prior reviews that have identified a set of factors (at the social-environmental level, at the school level, and at the student level) steering students' decisions to pursue STEM education or not (van den Hurk et al., 2019).

While most reviews to date have evaluated the effect of interventions on increasing interest in STEM, only a few have explored of the way the methodological and theoretical approaches have been applied and combined (i.e., Rosenzweig and Wigfield, 2016; Prieto-Rodriguez et al., 2020). We attempt to bridge this research gap by bringing together a scoping review of publications evaluating interventions or initiatives to increase young people's STEM participation, with particular attention on those that target girls. In addition, in alignment with previous reviews (Rosenzweig and Wigfield, 2016; van den Hurk et al., 2019), we will identify through a selection of interventions what type of school and social-environment level strategies have been used in order to tackle various personal factors influencing young people's interest in STEM pathways. Providing policymakers and educational practitioners with better information about the characteristics of effective STEM initiatives may help increase engagement and participation in STEM. It will also enable policymakers and practitioners to select the type of initiative that best suits their particular needs and interests (Australian Education Council, 2019).

Consistent with previous research (Rosenzweig and Wigfield, 2016; van den Hurk et al., 2019; Prieto-Rodriguez et al., 2020), this review focuses on middle and high school, the educational stages where decisions about academic and career pathways take place. Therefore, the primary goal of this review is to examine the main features and strategies deployed by a selection of intervention studies in order to inform the design of future initiatives to increase girls' interest in STEM. The questions this review will address are as follows:

- **R.Q.1.** What are the characteristics of research (interventions) aimed at increasing interest in STEM subjects and/or careers and reduce the underrepresentation of girls in STEM?
- **R.Q.2.** What methods have been used to measure the effectiveness of these interventions?
- **R.Q.3.** Which intervention studies are the best examples for inspiring and guiding the design of future interventions?

A systematic search of the literature in five databases and additional search strategies were used to respond to the three afore-mentioned research questions. More specifically, the ultimate goal of this review is to examine existing methodological and theoretical gaps in the different identified interventions. This could provide guidelines and recommendations for the design and development of future interdisciplinary intervention strategies.

Materials and methods

Design

A scoping review methodology was used (Arksey and O'Malley, 2005) in this study. This approach is particularly useful when, as in this review, researchers are interested in identifying the scope and extent of published research on a particular research topic and in examining how this research has been carried out (Arksey and O'Malley, 2005; Grant and Booth, 2009; Munn et al., 2018). Since the purpose of our review was to identify knowledge gaps and scope a body of literature about STEM intervention studies to raise preferably girls' interest in STEM—rather than producing a synthesized answer to a particular question— we chose to carry out a scoping review instead of a systematic review (Munn et al., 2018). Thus, the scoping review was the most suitable systematic reviewing methodology to determine the coverage of the wide range of literature that evaluates STEM interventions for secondary students, to provide a detailed overview of this literature, and to identify the most important literature gaps. In the conduction and reporting of this review, we adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for scoping reviews (PRISMAScR). These guidelines have been outlined by Tricco et al. (2018). These guidelines, outlined by Tricco et al. (2018), include, among others: specifying the characteristics of the sources of evidence used as eligibility criteria and providing a rationale; describing all information sources in the search as well as the date of the most recent search; presenting the complete electronic search strategy for at least one database, including any limits used; and describing the process for selecting sources of evidence included in the scoping review. When preparing the methods section and the remaining sections of the review, we ensured that all of the aforementioned guidelines of the PRISMA-ScR were followed.

Search strategy

A systematic search of empirical literature published in English between 1998 and 2019 was carried out in the following five databases: APA PsycNET, ERIC, ProQuest, Scopus, and Web of Science. These databases were selected because of their broad coverage of literature on science and technology, education, behavioral sciences and mental health, social sciences, the arts, and the humanities. The searches were carried out in the title and abstract fields using search terms associated with the following four concepts: interventions, STEM studies and professions, outcomes, and gender. The search query used, developed with the assistance of an information scientist from the Universitat Oberta de Catalunya, is shown in Table 1.

TABLE 1 Search.

Concept	Search terms (in title or abstract)
Intervention	Program* OR Interven* OR Initiative* OR Strateg* OR Seminar* OR Workshop* OR Course* OR Session*
STEM studies and professions	STEM OR Math* OR Science* OR Scient* OR Engineer* OR Technolog* OR "Technical stud*" OR career* OR "Technical career*" OR "Technical occupation*" OR "Technical subject*" OR "Scientific career*" OR "Scientific stud*" OR "Scientific occupation*" OR "Scientific subject*"
Positive outcomes	Interest* OR Engag* OR Motivat* OR Perform* OR Score* OR Grade* OR Abilit* OR Achiev* OR Choice* OR Selection OR Self-efficacy OR "Self-competence*" OR "Self-perception* of abilit*" OR "Sense of belonging" OR Stereotyp* OR Attitude* OR Participat* OR Involv* OR Capab* OR Encourag* OR Increas* OR Aspiration*OR "Self-concept*"
Gender	Gender OR Girl* OR Female* OR Woman OR Women OR Sex

10.3389/fpsyg.2022.954996

TABLE 2 Influential journals in the social and behavioral sciences.

Journal name	Impact factor (Journal citation reports, 2021)
American Psychologist	16.358, Q1, Psychology, Multidisciplinary, 4/147
Annual Review of Psychology	27.782, Q1, Psychology, 1/79
Developmental Psychology	4.497, Q2, Psychology, Developmental, 19/78
Educational Psychology Review	8.240, Q1, Psychology, Educational, 1/61
Educational Research	2.968, Q2, Education & Educational Research, 90/267
International Journal of Science Education	2.518, Education & Educational Research, 127/267
Journal of Applied Developmental Psychology	3.280, Q2, Psychology, Developmental, 33/78
Journal of Educational Psychology	6.856, Q1, Psychology, Educational, 4/61
Journal of Experimental Child Psychology	2.547, Q3, Psychology, Developmental, 47/78
Journal of Personality and Social Psychology	8.460, Q1, Psychology, Social, 3/65
Personality and Social Psychology Bulletin	4.560, Q2, Psychology, Social, 18/65
Perspectives on Psychological Science	11.621, Q1, Psychology, Multidisciplinary, 6/147
Psychological Bulletin	23.027, Q1, Psychology, 3/79
Psychological Science	10.172, Q1, Psychology, Multidisciplinary, 9/147
Psychology of Women Quarterly	4.292, Q1, Psychology, Multidisciplinary, 33/147
Review of Educational Research	13.551, Q1, Education & Educational Research, 1/267
Science	63.798, Q1, Multidisciplinary Sciences, 2/73
Sex Roles	3.812, Q2, Psychology, Developmental, 25/78
Social Psychological and Personality Science	5.316, Q1, Psychology, Social, 12/65
Social Science Quarterly	1.781, Q3, Political Science, 106/187

Three additional search strategies were used to complement the database search. First, 20 influential journals in the social and behavioral sciences were hand searched (see Table 2): Second, lists of publications from influential authors in the field were reviewed for studies not identified in the database search. Third, citation searching was carried out by scanning the references cited by key articles.

Inclusion and exclusion criteria

In order to be included in the review, publications had to meet four inclusion criteria. First, they needed to report empirical research evaluating interventions for promoting the participation of secondary school students in STEM fields. Second, they needed to describe the aims, participants, and context of the intervention and provide a succinct description of its implementation. Third, they needed to evaluate the effectiveness of the intervention through clearly defined and operationalized outcomes using either quantitative, qualitative, or mixed methods. Fourth, they needed to be in English and published between 1998 and 2019. All types of publications were included, including journal articles, books, book chapters, and dissertations. Studies in which the participants were the parents or secondary school teachers of students were also included. Non-empirical articles, such as systematic reviews, editorials, or commentaries, were excluded.

Study selection

The publications retrieved from the databases and those identified through the complementary search strategies were imported into the EPPI Reviewer software, which was used to facilitate the study selection. The selection was carried out in two phases. In Phase 1, two researchers independently screened a random sample of 10% ($n = 4.017$) of the articles. Each reviewer screened the same number of articles and disagreements between the two researchers were resolved through discussion with the involvement of a third reviewer, when necessary. Inter-rater agreement was high (Kappa = 0.825). The remaining articles were divided between the two researchers. In Phase 2, the full text of the eligible publications was independently reviewed by the same two researchers. Disagreements in this phase were again resolved by consensus. The percentage of discrepancies between the two researchers during the screening and eligibility phases was similar.

Data extraction and synthesis

The steps described by Arksey and O'Malley (2005) were followed during the data extraction and synthesis. First, data from the included publications was extracted in Excel using a standardized tool. The following information was collected: publication metadata (i.e., publication year and type affiliation), intervention characteristics (i.e., purpose, participants' profile, setting, and theory motivating the intervention), focus of the evaluation, and methodological features of the study. Two researchers independently performed the data extraction. Disagreements between the researchers were discussed among the members of the research team until a consensus was reached. Second, once the extraction was completed, summary tables were generated to chart the extracted data and compare between intervention study types. This comparison allowed the researchers to identify patterns across the included studies and generate a narrative account of the results. At this point, researchers returned to the original publications several times to ensure that the summaries were supported by the data.

Results

The database search generated 52.622 publications, of which 52.502 were identified through database searching, while 120 were identified using the additional search strategies. After removing duplicate publications and assessing eligibility, 215 publications were included. Figure 1 shows the Preferred Reporting Items for Systematic Reviews and a Meta-Analyses (PRISMA) flow diagram of the review process.

R.Q.1. Characteristics of the publications and the interventions

Most of the publications (*n* = 180) were published between 2010 and 2019 (see Table 3). The majority of publications were journal articles (*n* = 147), followed by conference proceedings (*n* = 34), and dissertations (*n* = 25).

The main disciplines of corresponding authors were psychology (*n* = 63), STEM fields with a high technological component (i.e., engineering, computer science, and information technologies, *n* = 48), education (*n* = 47), and STEM fields with a high scientific component (*n* = 45) including mathematics, the physical sciences, medicine, pediatrics, chemistry, biomedicine, agriculture, and biology.

As shown in Table 4, the majority of intervention studies (*n* = 95) focused on scientific STEM fields, mainly biology, the physical sciences, mathematics, and chemistry, while others *n* = 45) considered STEM fields with a high technological component (i.e., computer science or engineering). Many of the interventions (*n* = 71) did not identify any STEM field. Only three included a combination of arts and STEM competences (STEAM).

In terms of geographical location, the majority (*n* = 149) of the studies were conducted in North America, particularly the United States. Likewise, most of the interventions (*n* = 176) were conducted exclusively with students.

In addition, while most of the intervention studies (*n* = 146) targeted both genders, 66 focused solely on female students. Furthermore, 85 of the intervention studies had a short duration (1 or 2 h session, 1 day, or 1 week), and 72 had a mid-term duration (between 2 weeks and 4 months). Interestingly, 58 of the intervention studies were long term, lasting more than 14 weeks.

The vast majority of interventions were organized in the context of regular classroom activities (*n* = 119) and were hands-on (*n* = 100). Moreover, the greatest part of the extracurricular activities consisted of summer camps (*n* = 33), workshops and presentations (*n* = 22), and afterschool activities (*n* = 20). Finally, while the majority of intervention studies explicitly drew on theoretical foundations (*n* = 127), 88 did not declare any particular theoretical approach.

R.Q.2. Methodology of the interventions

A great number of the intervention studies (see Table 5) were exclusively based on a quantitative methodology (*n* = 141), whereas the rest on a mixed methods approach (*n* = 73). Only one study had a qualitative nature, which relied on interviews for collecting and analyzing data. Most of the 141 quantitative studies, had a quasi-experimental design (73), and 39 had an experimental design. Interestingly, a high proportion of the quasi-experimental studies (*n* = 47) had a pre-experimental design. Among the experimental studies, most of them applied a single factor inter-subject design (*n* = 27). The methods employed included self-reported surveys (*n* = 77), achievement tests (*n* = 8), and grades (*n* = 1), or combined various quantitative methods (*n* = 55).

Curiously (see Table 6), the majority of intervention designs that used a mixed methods approach were convergent designs (*n* = 63). Designs used in the quantitative strand of mixed methods studies were mostly quasi-experimental (*n* = 57). The majority of the quasi-experimental studies applied a pre-experimental design (*n* = 43). Among the experimental studies (*n* = 9), most of them (*n* = 7) applied a single factor inter-subject design. The designs used in the qualitative strand of mixed methods studies were largely qualitative description (*n* = 45).

Methods used in the quantitative strand of the mixed methods studies were mostly based on self-reported surveys (*n* = 43), whereas the qualitative strand in the mixed methods studies were mainly based on interviews (*n* = 16), open-ended questions (*n* = 15), or on various methods used simultaneously (*n* = 34).

As shown in Table 7, 30 of these interventions drew on Expectancy-Value theories (i.e., Victor, 2005; Eccles, 2009), 22 on Constructivist Learning Theories, and 17 on Social Learning Theory (Bandura's Self-Efficacy theory). Only three were inspired by feminist theories and feminist pedagogy, critical mass, and intersectional theories. Interestingly, 27 of the intervention studies relied on more than one motivation theory (competence-related beliefs, expectancy-value, attributions, or theories of intelligence).

Curiously, some of the intervention studies that did not draw on any particular theory included learning and the acquisition of knowledge on specific STEM content. For instance, the study by

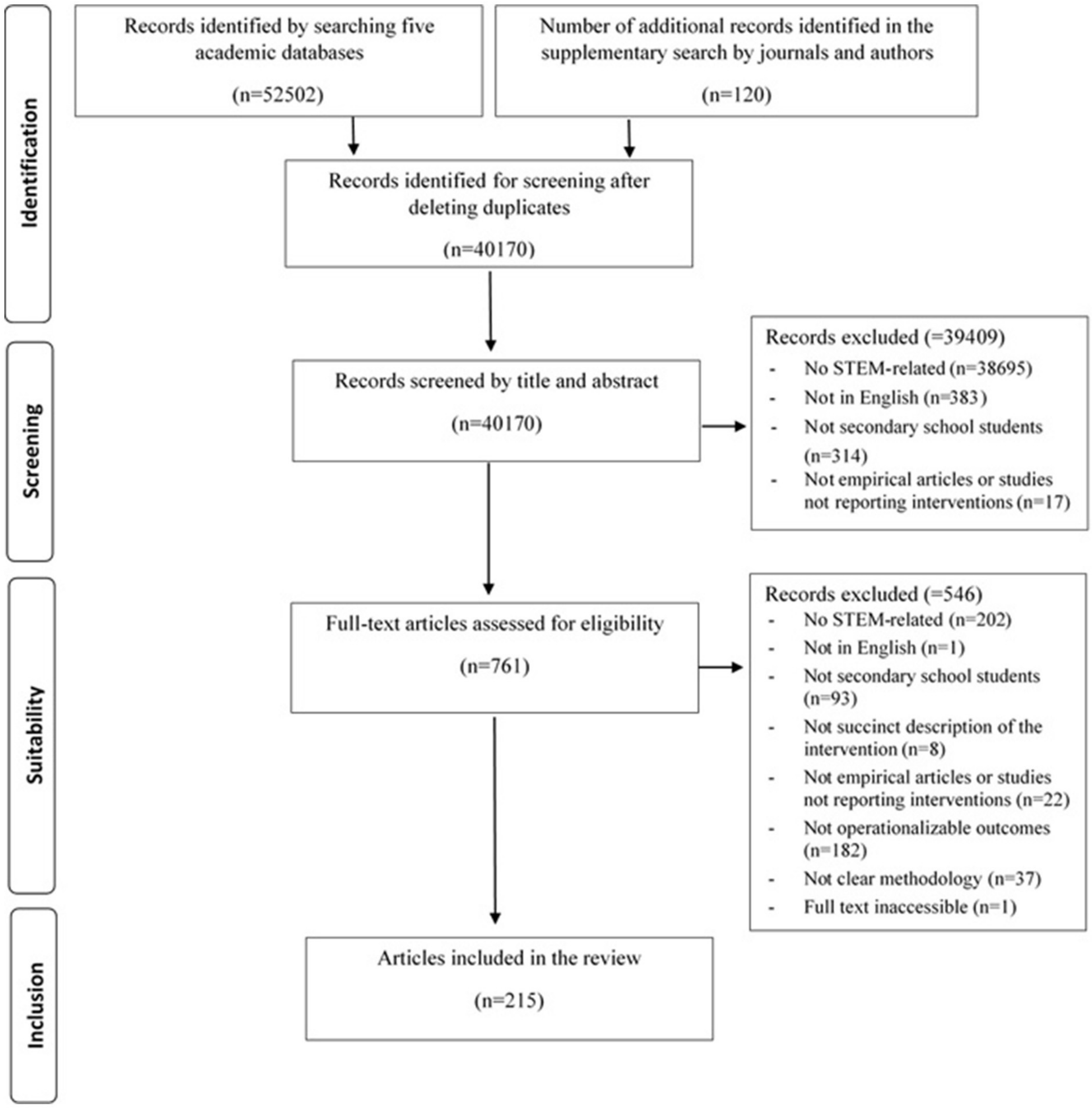

FIGURE 1
Flow-chart diagram illustrating inclusion and exclusion of articles during search process.

Tarng et al. (2011) provided the basic concepts of synchrotron light sources, while other interventions focused on students' attitudes and interest in STEM subjects and careers (Christensen et al., 2014, 2015; Christensen and Knezek, 2017; Acuña et al., 2018) or STEM performance (Brown and Brown, 2019; Jordaan and Tavenga, 2019; Mostoli et al., 2019).

In line with the theoretical foundations of interventions, 74 measured various motivational constructs as outcomes to evaluate effectiveness: self-concept of ability, self-efficacy, perceived utility value of STEM subjects, interest in pursuing STEM studies, intrinsic value of STEM, and attainment. Only 14 intervention studies exclusively focused on achievement, whereas 75 interventions combined the use of different motivational and achievement-related constructs.

TABLE 3 Characteristics of the publications.

	n	%
Year of publication		
2000–2009	35	16.3
2010–2019	180	83.7
Type of publication		
Journal article	147	68.4
Dissertation	25	11.6
Book chapter	7	3.3
Conference proceedings	34	15.8
Working paper or report	2	0.9
Corresponding author discipline		
Arts & humanities	6	2.8
Education	47	21.9
Psychology	63	29.3
Science (STEM)	45	20.9
Technology (STEM)	48	22.3
Social sciences	6	2.8

Categories are mutually exclusive. The percentages are calculated relative to the number of articles that included information on this feature.

TABLE 4 Characteristics of the interventions.

	n	%
STEM field		
Science	95	44.2
Technology	45	20.9
Various STEM disciplines	71	33.0
STEAM	4	1.9
Geographical location		
Africa	3	1.4
Asia	14	6.5
Europe	43	20.0
Latin America	1	0.5
Latin America & Africa	1	0.5
North America	149	69.3
Oceania	4	1.9
Target people		
Only students	176	81.9
Students & parents	8	3.7
Students & teachers	21	9.8
Students & mentors (peers or professionals)	5	2.3
Combination of several previous categories	5	2.3
Gender		
Boys and girls	146	67.9
Only girls	66	30.7
Combination of the previous	3	1.4
Duration		
Short-Term	85	39.5
Mid-Term	72	33.5
Long-Term	58	27.0
Type of activity		
Ordinary classroom activity	119	55.3
Extracurricular activity	91	42.3
Both	5	2.3
Ordinary classroom activity		
Hands-On classroom activities	100	84.0
Workshops	5	4.2
Laboratory experiments	4	3.4
Excursions	4	3.4
Games	4	3.4
Counseling sessions	2	1.7
Extracurricular activities		
Summer camps	33	36.3
Competitions	2	2.2
Out of school/after school activities	24	26.4
University camps	10	11.0
Workshops, Presentations	22	24.2
Theory feeding the intervention		
Yes	127	59.1
No	88	40.9

R.Q.3. Classification of outstanding interventions

A total of 25 intervention studies were selected according to various attributes of the intervention (context of the intervention, gender of target people, and main findings), and the intervention evaluation (main purpose, theory feeding the interventions, and evaluation method). That is, the interventions met the following criteria (see Table 7).

- Interventions drawing on one or more theoretical approach since this information serves as an indicator of the operationalization of theory in practice (Lazowski and Hulleman, 2016; Rosenzweig and Wigfield, 2016).
- The methodological approach was based on a quantitative or mixed methods design, since it provides further insights about the tools used to conduct the evaluation of the intervention (Kolne and Lindsay, 2020; Prieto-Rodriguez et al., 2020).
- The design of the intervention included an experimental or a quasi-experimental design, as it informs about the quality of the evaluation of the intervention (Rosenzweig and Wigfield, 2016; van den Hurk et al., 2019; Prieto-Rodriguez et al., 2020).
- The duration of the intervention lasted either more than 2 weeks (a mid-term) or several months or years (long-term), since this informs about the sustainability and long-term strategy of the intervention (Prieto-Rodriguez et al., 2020).

10.3389/fpsyg.2022.954996

TABLE 5 Methodology of the intervention studies with a quantitative approach.

	n	%
Method		
Quantitative	141	65.6
Qualitative	1	0.5
Mixed methods	73	34.0
Type of design in the quantitative studies		
Experimental	46	32.6
Quasi-Experimental	67	47.5
Cohort	22	15.6
Cross-Sectional	4	2.8
Various designs	2	1.4
Type of experimental design		
Single factor inter-subject design: 2 groups with pre- & post-test or only post	27	58.7
Single factor inter-subject design: multi-group with pre- & post-test or only post-test	18	39.1
Single factor within-subject design: single group with pre- & post-test	1	2.2
Type of quasi-experimental		
Pre-Experimental design: single group with pre- &post-test	47	70.1
Quasi-Experimental design, with non-equivalent control group with pre-post-measurements	14	20.9
Quasi-Experimental design: interrupted time series, simple (1 group, pre- & post-test)	4	6.0
Quasi-Experimental design: interrupted time series, with non-equivalent control group (pre & post-test)	2	3.0
Type of method used in the quantitative studies		
Achievement tests	8	5.7
Self-Reported surveys	77	54.6
Grades	1	0.7
Various methods	55	39.0

Categories are not mutually exclusive. The percentages are calculated relative to the number of articles that included information on this feature.

- The purpose and/or research questions included a gender perspective, since they provide a clear commitment of the intervention to tackle the dearth of women in STEM pathways (Tannenbaum et al., 2019).

The majority of the selected exemplars applied to the field of science (8/25), computer science (6/25), math (5/25), and STEM (4/25). Only one intervention applied to the field of engineering and another one to physical science.

Most of the selected interventions attempted to improve girls' and students' STEM motivation, performance, skills or competences, and self-perceptions (self-efficacy). To achieve the purpose nine of the selected interventions applied a STEM training strategy through courses, workshops, and summer camps (9/25). Brock's quasi-experimental study developed in the context of a math workshop (Brock, 2017) helped to reduce the gender gap in math achievement between advanced

TABLE 6 Methodology of the intervention studies with a mixed-method approach.

	n	%
Type of design of the mixed methods studies		
Convergent	63	86.3
Explanatory sequential	10	13.7
Type of design of the quantitative strand of the mixed methods studies		
Experimental	9	12.3
Quasi-Experimental	56	76.7
Cohort	7	9.6
Various designs	1	1.4
Type of experimental design of the mixed methods studies		
Single factor inter-subject design: 2 groups with pre- & post-test or only post	7	77.8
Single factor inter-subject design: multi-group with pre- & post-test or only post	2	22.2
Type of quasi-experimental design of the mixed methods studies		
Pre-Experimental design: single group with pre-post test	43	76.8
Quasi-Experimental design, with non-equivalent control group with pre-post-test	8	14.3
Quasi-Experimental design: interrupted time series, simple (1 group, pre- & post-test)	3	5.4
Quasi-experimental design: interrupted time series, with non-equivalent control group (Pre- & Post-test)	2	3.6
Type of design qualitative strand of the mixed methods studies		
Case study	1	1.4
Content analysis	5	6.8
Ethnography	1	1.4
Grounded theory	1	1.4
Phenomenology	1	1.4
Qualitative description	45	61.6
Non-reported design	19	26.0
Type of method used in the quantitative strand of the mixed methods studies		
Achievement tests	5	6.8
Self-Reported surveys	43	58.9
Various methods	25	34.2
Type of method used in the qualitative strand of the mixed methods studies		
Open-Ended questions	15	20.5
Interviews	16	21.9
Focus groups	5	6.8
Journal entries	2	2.7
Observations	1	1.4
Various qualitative methods	34	46.6

Categories are not mutually exclusive. The percentages are calculated relative to the number of articles that included information on this feature.

TABLE 7 Theories and outcomes associated with the interventions.

	n	%
Theories feeding the interventions		
Expectancy-Value theory	30	14.0
Social learning theory (self-efficacy)	17	7.9
Social role theory	2	0.9
Role model theory	5	2.3
Sociocultural learning theory (Vygotsky)	4	1.9
Feminist theories	3	1.4
Self-Determination theory	2	0.9
Constructivist learning theories	22	10.2
Career development (Holland) theory	1	0.5
Theories of intelligence	7	3.3
Identity theories	3	1.4
Self-Affirmation theory	1	0.5
Theories of emotion	1	0.5
Attribution theory	1	0.5
Stereotype threat theory	1	0.5
Various theories	27	12.6
Outcomes		
Achievement	14	6.5
Motivation	74	34.4
Gender stereotypes	2	0.9
Achievement & motivation	74	34.4
Achievement & emotion	2	0.9
Achievement, motivation, & identity	3	1.4
Motivation & identity	9	4.2
Motivation & gender stereotypes	19	8.8
Motivation & emotion	5	2.3
Achievement, motivation, gender stereotypes, & emotion	4	1.9
Achievement, motivation, gender stereotypes, identity, & emotion	1	0.5
Achievement, motivation, & gender stereotypes	4	1.9
Motivation, gender stereotypes, & identity	3	1.4
Achievement, motivation, & emotion	3	1.4
Achievement, motivation, gender stereotypes, & identity	1	0.5

Categories are mutually exclusive. The percentages are calculated relative to the number of articles that included information on this feature.

students. Similarly, Isiksal and Askar's (2005) experimental study made use of two software programs and observed no significant gender differences in mathematics achievement and self-efficacy. Denner's (2007) quasi-experimental study based on a STEM training program focused on improving girls' computer skills revealed that girls improved their computer skills, knowledge about computers, and perceived social support. Girls' perception that boys do better than girls with computers was reduced. Likewise, in Hall-Lay's (2018) quasi-experimental study developed in the framework of STEM programs, students who participated in robotics programs scored significantly higher than students enrolled in other STEM-related programs. However, no gender differences were observed.

Paslov (2006) quasi-experimental study demonstrated that girls who participated in the program improved their self-efficacy and achievement in mathematics. Similarly, Scott et al. (2017) quasi-experimental study with a computer science preparatory course showed that girls from ethnic group's interest in computer science increased over time, despite their initial lack of interest. Male students showed higher interest and aspiration in computer science. There were no gender differences in course completion, but taking a course did not improve female students' likelihood of majoring in computer science. In Todd and Zvoch (2019a) experimental study, girls who participated in the summer camp scored higher in science efficacy and attitudes toward science than girls in the control group. No significant differences were observed in science interest and science identity. In a similar vein, Todd and Zvoch (2019b) quasi-experimental research showed that girls from affluent families participating in a summer science program increased their affinities in science over time. However, among girls from low income families their early gains in affinities in science diminished over time. Finally, Ziegler and Heller (2000) experimental study demonstrated how an attribution program in physics improved girls' performance as well as their motivation and self-related cognitions in physics.

Other two of the selected interventions included evaluations of single-sex and coeducational contexts (2/25). In this way, Drobnis (2010) quasi-experimental study consisting of a computer science summer course concluded that boys in the mixed gender group and girls in the only girls group had higher computer self-efficacy and higher gains in computer science scores than girls in the mixed gender group. Interestingly, in Schilling and Pinnell (2019) experimental study by allowing participants to explore engineering in a positive environment and encouraging them to work through challenges, participants built confidence in engineering.

Five of the selected interventions developed changes in the curriculum and pedagogy (5/25). In Cantley et al. (2017) quasi-experimental research the pedagogical mathematics tool used increased girls' interest and enjoyment of mathematics, but it had no significant change in boys' attitudes. Additionally, no significant gender differences in pre-intervention enjoyment scores were observed. Chiu's (2011) experimental study focused on the recognition of women in science and men in humanities, awareness of academic gender stereotypes, and development of unique selves when learning science. In comparison to the control group, students in the experimental group did not experience any change in their attitudes toward learning science. Boys in both the experimental and control groups increased the value attached to learning science throughout the pre-test phase. However, in the experimental group girls' value exclusively increased during the post-test phase.

Mayberry's (2015) quasi-experimental research found that female-oriented curriculum improved girls' interest and confidence in STEM careers. Similarly, McHugh et al. (2018) quasi-experimental study examined the influence of incorporating mathematical skills into the curriculum as a complement to science content on students' achievement and attitudes toward science. Females outperformed males in science, but there were no significant differences in achievement between students from high and low-needs schools. Remarkably, Werner's (2017) quasi-experimental study deployed a female-oriented pedagogy that increased female science students' positive attitude toward self-concept in science, enjoyment of science, and perception of science teacher. Female science students also developed less anxiety about science.

Furthermore, five of the selected interventions revolved around the use of female role models and mentoring strategies (5/25). In this regard, Stake's (2006) quasi-experimental research demonstrated that boys' stereotypes toward women in science could be changed by exposing them to female role models and mentors, as well as to positive information about girls' and women's science abilities. In addition, Stoeger et al. (2013) experimental study confirmed the advantages of 1-year female mentoring program in increasing girls' STEM activity, self-assessment of knowledge of STEM topics, self-assessment of knowledge about STEM-related university studies and jobs, confidence in one's own STEM abilities, self-assessment of STEM competences, and academic elective intentions.

Likewise, Denner's et al. (2012) quasi-experimental investigation observed that an after-school summer program with professionals serving as virtual mentors increased girls' computing career goals, expectations for success with computing, the value they placed on computing and computing-related jobs, and their perceived parental support. Strikingly, Good et al. (2003) experimental study significantly boosted the performance of girls, minority, and low-income students by addressing the psychologically threatening nature of the math assessments. Students were mentored by college students who encouraged them to view intelligence as malleable or to attribute academic difficulties to the novelty of the educational setting. Finally, Wilson's (2019) quasi-experimental research focused on analyzing the effectiveness of a STEM program on students' and teachers' efficacy and attitudes toward STEM. While girls' STEM confidence increased over time, no relationship between teachers' preparation and self-efficacy in STEM and students' STEM confidence was observed. Additionally, girls and students from ethnic groups felt as confident as boys and students from non-ethnic groups with their ability to learn with STEM resources and equipment.

Two of the remaining selected interventions were counseling-oriented (2/25). That is, experimental studies by Falco et al. (2010) and Falco and Summers (2019) observed improvements in girls' career decision self-efficacy and STEM self-efficacy as well as in students' motivation, value, enjoyment, and confidence in mathematics after the counseling sessions.

Finally, only two of the selected interventions focused on increasing parents' engagement (2/25). Heddy's (2014) experimental study demonstrated that combining a Teaching for Transformative Experience in Science (TTES) and a parent involvement (PI) intervention potentially ameliorated the reduction in girls' STEM motivation. In Hyde et al. (2017) experimental research, hypothetical responses of mothers to their children's usefulness of math and science classes increased adolescents' perception of math ability in seventh grade. Those responses also positively predicted adolescents' STEM interest in 10th grade.

In conclusion, the majority of the selected studies targeted personal factors through changes at the school level ($n = 18$), the delivery of STEM training ($n = 9$), modifications in pedagogy and curriculum of STEM content ($n = 5$), the promotion of single-sex and coeducational school context ($n = 2$), or the use of counseling sessions ($n = 2$). The rest of the intervention studies targeted changes in personal factors through a series of strategies at the environmental level ($n = 7$). Whereas, most of these used female role models ($n = 5$), the rest promoted parental engagement ($n = 2$).

Discussion

This article provides a scoping review of interdisciplinary interventions designed to increase young people's interest in STEM and more particularly girls' interest in those STEM fields where women are highly underrepresented like engineering and computer science (Cheryan et al., 2013; UNESCO, 2018; Sáinz, 2020). The findings of the present study expand current knowledge on how to measure the impact and effectiveness of interventions designed to increase young people's interest in STEM. It responds to all research questions through an analysis of the characteristics of intervention studies aiming at raising young people's interest in STEM (RQ1) and of the methods used to measure the effectiveness of the interventions (RQ2). In addition, 25 exemplar intervention studies were selected to illustrate different approaches to addressing the topic, which can be a source of inspiration for the design and implementation of future intervention studies (RQ3).

In line with previous reviews and with research conducted in some of the studies included in our review, the present study confirms the need for designing interventions that lead to changes at family and school levels. In this regard, this review provides scholars, practitioners, policy-makers, and the general public with practical evidence of intervention studies that have fully or partially succeeded through the use of different strategies at the environmental and school levels in changing various personal aspects involved in shaping girls' and young people's interest in STEM.

However, and in comparison to previous reviews in this area, the present study provides a broader scope in the type of included publications. On the one hand, the 20-year time framework of this review is wider-ranging than earlier reviews. On the other hand, 25 dissertations were part of the initial review, but only five were included in our selection of exemplars. One of the strengths of the present review is a recognition of the theoretical foundations inspiring the methodological designs involved in the evaluation of the effectiveness of the interventions. Many of the identified characteristics of both publications and interventions deal with the methodology used to conduct the different studies, including the type of mixed method design, the typology of quantitative or qualitative research methods and techniques, as well as the kind of quantitative design used to collect, analyse, and interpret the collected data. All these features have not been comprehensively taken into consideration by earlier research on the effectiveness/impact of the evaluation of intervention studies. This is especially true in the case of the methodological design associated with interventions that have applied a mixed method approach.

Measurement of the effectiveness of interventions to reduce the gender gap in STEM

A surprising amount of intervention studies did not have a stated theoretical foundation. This calls into question both the validity of designs and outcomes along with the extent to which the components of intervention evaluation have been fully comprehended (Liben and Coyle, 2014; Lazowski and Hulleman, 2016; Rosenzweig and Wigfield, 2016). In this review, motivation theories like expectancy-value, social learning, theories of intelligence, as well as constructivist theory shape most of the researchers' understanding and design of the studies addressing outcomes involved in shaping young people's interest in STEM pathways (Rosenzweig and Wigfield, 2016). Most of the studies included in this review use various constructs to measure issues associated with STEM achievement, STEM motivation, gender stereotypes about STEM careers, STEM identity, and emotional response toward STEM fields (Rosenzweig and Wigfield, 2016; van den Hurk et al., 2019).

These studies describe the different intervention practices and strategies used to target these constructs at the social-environmental (i.e., the use of female role models) and school levels (i.e., the inclusion of STEM training, changes in STEM pedagogy, the use of counseling sessions, or the influence of single-sex & co-educational contexts). Interestingly and in line with Rosenzweig and Wigfield (2016) findings, a great number of the studies informed about changes in various personal factors. However, most of them did not develop a proper theory of change explaining the psychological, social, environmental processes involved in this change.

Although a qualitative appraisal of interventions was not applicable due to the high number of publications covered, the identified interventions applied quantitative research methods. Random assignment was not possible for some of the interventions, due to the ethical constraints associated with educational fieldwork (i.e., Brock, 2017). That is, the random assignment of participants to experimental groups was not possible because the groups were already predefined previously to the study in classes or grade levels, and researchers could not reorganize them to meet the needs of the research (Mertler, 2016; Brown, 2019). Although several of the intervention studies applied a quasi-experimental design, a considerable number of them did not include a control group.

Interestingly, several studies combined qualitative and quantitative research methods. This reinforces the importance of using both methodological approaches when tackling complex phenomena, such as the underrepresentation of women in STEM. In contrast to previous systematic reviews (Rosenzweig and Wigfield, 2016; van den Hurk et al., 2019; Prieto-Rodriguez et al., 2020), studies using a mixed-method approach for measuring the effectiveness of interventions were included. Most of these studies used a convergent design (Creswell and Plano Clark, 2018) in order to gain complementary insights from the quantitative and qualitative findings regarding how, why, and under what conditions a particular intervention was successful or not in raising young people's interest in STEM.

Our scoping review highlights the need to design and include qualitative methods to evaluate the impact and effectiveness of interventions. This will enable a richer evaluation of the true influence on a targeted group. For instance, in Werner's (2017) study whereas quantitative data gauged the degree of change in six facets of attitude among female students, interviews were also used to gain deeper insights into student perspectives.

Exemplary interventions that address the gender gap in STEM participation

A selection of 25 interventions of various types has been provided in this review, affording a useful reference point for the design of future research that can distinguish between short-term and long-term effects of interventions.

Since our research had a clear focus on studies evaluating intervention effectiveness, quantitative methods were used as an inclusion criterion for the exemplar studies as long as they allow researchers to accurately assess effectiveness, measuring changes in outcomes before and after the intervention (Mertler, 2016). Additionally, we recognize the value of qualitative methods, particularly when combined with quantitative methods in mixed methods designs. Especially when they aim at achieving additional evaluation objectives, such as determining the acceptability and feasibility of the intervention or determining how participants perceived the effectiveness of the intervention (Fetters and Molina-Azorin, 2020).

In line with van den Hurk et al. (2019) review, most of the identified intervention studies targeted various personal level factors linked to young people's interest in STEM, such as attitudes toward STEM learning, STEM self-efficacy, and self-perception of competence, or STEM achievement. Whereas, some of the reviewed studies employed strategies that modified environmental factors such as the increase of parental engagement or the use of female role models, other studies developed strategies that changed school level factors such as the inclusion of pedagogies in the teaching of STEM subjects, or the promotion of co-educational classrooms. This is another evidence of the complexity of the phenomenon associated with women's under-representation in several STEM pathways. In this regard and as shown in previous systematic reviews (Rosenzweig and Wigfield, 2016; Prieto-Rodriguez et al., 2020), most interventions used several constructs when raising girls' and young people's interest in STEM.

Incidentally, several of the selected interventions targeted personal factors, including self-competence beliefs identity formation sense of belonging, attribution of STEM-related success or failure, interest in STEM, and achievement, by modifying various aspects either at school (such as changes in STEM curriculum, the content of STEM training programmes, and teaching strategies) or at social-environment level (such as valuing women in science, or making visible female role models). Curiously, in some studies personal factors also tend moderate the effect that other factors of the intervention have on different indicators of STEM education. For instance, in Brock's study the efficacy of the intervention to increase math achievement was moderated by the students' sense of belonging. Similarly, in Todd and Zvoch (2019b) study, positive attitudes alone were not enough to increase girls' persistence in STEM. Self-efficacy and identities in science were also needed.

Interestingly, most of the selected interventions achieved their aims, whether it was closing the gender gap in STEM achievement or increasing interest in STEM particularly among female students. These interventions accomplished their goals by increasing the value placed on women in science improving female students' self-perception of STEM and encouraging them to pursue and persist with STEM, increasing students' self-efficacy in STEM, raising girls' interest in STEM by changing stereotypical images of computer science, retraining attribution toward physics, and increasing science affinities among female students.

Despite their qualities, several of the selected interventions demonstrated only partial effectiveness. For instance, Werner's (2017) intervention using female-oriented teaching strategies was insufficient to increase female science students' motivation. In a similar vein, Todd and Zvoch's (2019) intervention demonstrated that positive attitudes toward identity and self-efficacy among girls were not enough to increase persistence in STEM. This finding suggests that the impact of interventions addressing persistence in STEM should be measured longitudinally. Similarly, despite increasing interest in computer science among female students of color, the intervention by Scott et al. (2017) was unable to fully close the gender gap in interests and aspirations. This last finding confirms the importance of measuring long-term effects of the interventions implemented, especially in fields like computer science, where the participation of women is really scarce.

Moreover, a proper design of the interventions involves uunderstanding and anticipating the dynamics between early subjective STEM experiences and social and/or environmental challenges to STEM education (Schoon, 2001). In fact, in Schilling and Pinnell's study there were few opportunities for female participants to fully participate in engineering-related activities in the mixed gender group (i.e., they were given tasks associated with feminine roles, such as taking notes for the group), The inclusion of moderating factors also introduced complexity in assessing the real effect of the interventions. In Mayberry's (2015) study, after the changes made in the curriculum of science parental and educational background had no effect of on girls' interest, confidence, desire to learn about STEM and motivation to pursue STEM careers.

Many of the selected studies measured the effect of the intervention on more than a single construct. However, the intervention did not result in significant effects of the intervention on all the considered constructs. In Denner's et al. (2012) study, whereas the intervention with female virtual mentors had effect on some of the constructs under research (i.e., interest in computing jobs, confidence in computers, computer use, or perceived support from parents), girls' interest in problem-solving, endorsement of gender stereotypes, and perceived support from peers and teachers did not change. In Chiu's (2011) study the intervention focused on changing STEM pedagogy had no effects on students' attitudes toward science who participated in the experimental group.

Limitations

This study has several limitations. First, despite the systematic and rigorous search strategy, we might have missed relevant intervention studies published in books or dissertations not indexed in the selected databases. Second, in accordance with the disciplinary background of the authors of this review, most of the journals included as part of the complementary search strategies were from the social, educational, and behavioral sciences, while journals from other STEM-related domains, such as engineering or from disciplines like human resource management, or economics were not used in the hand search strategy. Nonetheless, this does not imply that research from these fields was omitted from the review, as no discipline-specific database search terms or exclusion criteria were used. Third, consistently with the scoping review methodology (Pham et al., 2014), the balance between breadth and depth of analysis was a challenge in our review. Since our selection criteria were broader than in previous systematic reviews, we could not provide a detailed appraisal of each of the reviewed interventions individually because of the large number of included publications. Fourth, the broad scope of the review also limited the depth of analysis in the appraisal of the interventions' effectiveness. Fourth, commensurate with the recommendation of privileging the comprehensive coverage of the literature over critical appraisal in scoping reviews, the methodological rigor of the included studies was not appraised.

Implications and recommendations

The present study highlights the important role that theory plays in the design and evaluation of the interventions to reduce the various gender gaps in STEM. However, and considering Yeager and Walton (2011) conclusions, the success of several of the interventions highly relies on societal and educational contexts. The theoretical and practical implications of this review on the effectiveness of interventions in the STEM field are numerous and can be a source of inspiration, both for the design of future interventions aimed at promoting the interest of girls in STEM fields and for their evaluation.

The high number of the identified interventions that focus on raising interest in STEM fields among girls and boys not only denotes a lack of STEM talent, but also the necessity of targeting both genders and incorporating a gender dimension in raising and retaining young people's interest in STEM. This issue not only relates to the underrepresentation of women in STEM, but it also involves aspects relating to the role that women play in STEM fields in particular and in society in general.

Similarly, since longitudinal evaluations do not necessarily demonstrate that interventions promoted long-term engagement in STEM (Prieto-Rodriguez et al., 2020), it is essential that this type of research measures a long-term engagement and not just simply a follow-up. Interestingly, in this review the benefits of combining qualitative and quantitative research methods in the design and implementation of interventions are highlighted. This implies future avenues for interdisciplinary collaboration in the design and implementation of interventions to raise women's interest in STEM.

In addition, given that most of the published interventions have been systematically conducted in the US (Liben and Coyle, 2014; Rosenzweig and Wigfield, 2016), more publications should be promoted tackling interventions implemented in different international countries with less international visibility of the initiatives and efforts implemented to fight against women's underrepresentation in STEM. These could also provide further cultural insights into how to improve the effectiveness of interventions addressing gender gaps in STEM pathways.

Interestingly, the combination of afterschool and within school activities could be an extraordinary way of increasing the likelihood of the interventions to improve women's attraction and retention in the STEM pathway over time. Many interventions took place in the classroom, through games, counseling sessions, and hands-on activities, and deployed various activities consistent with the STEM curriculum. Others were organized as extracurricular activities, such as summer camps and various afterschool activities, and incorporated diverse leisure activities, a highly important aspect in encouraging young people's vocational interests. These indicate the wide range of possibilities that both extracurricular and classroom activities have for the formulation of significant, innovative, and grounded interventions that can reach long-term goals.

From the results of the present review, we can assume that the inclusion of a gender perspective is a key element when analyzing the potential effect of gender issues in the way the interventions are designed and evaluated. Interventions must target both genders, as boys should also be exposed to interventions with female role models and learn more about contributions by women within STEM (Werner, 2017; González-Pérez et al., 2020). It is crucial that both girls and boys learn how existing gender roles about academic competences influence career choices. As noted by Good et al. (2003) and Stake (2006), the powerful influence of stereotype threat means that male bias may significantly limit the science performance of girls and women as well as their willingness to choose and persevere in STEM fields. Therefore, it is important that we understand the basis for negative attitudes toward women in STEM. As Harackiewicz and Priniski (2018) noted in their systematic review of interventions in higher education, if the design of interventions is not undertaken correctly, students of both genders may become disengaged and abandon STEM pathways.

Most of the interventions focused on increasing STEM interest, mainly in the fields of science, mathematics, and computer science, but a few of them attempted to increase interest in STEAM fields. This suggests that interventions combining the intersection between STEM competences and competences beyond STEM are needed. As Schilling and Pinnell (2019) study noted, some of the intervention studies failed to reflect the diversity of participants, as information about ethnicity and/or social origin was not gathered in many cases. Therefore, an intersectional analysis has not been conducted in most the intervention studies. For this reason, future design should incorporate various aspects relating to intersectionality and how it can help in properly addressing various gender gaps in STEM.

The duration of interventions might inform about their sustainability and impact of the intervention. Long-term exposure to STEM is needed to properly address the underrepresentation of women in STEM pathways. Ideally, more follow-up research is required in order to determine the long-term effect of interventions and the extent to which they are effective in increasing the number of high school and university STEM enrolments, as well as retain women and students with different characteristics in STEM pathways.

Future research should provide more rigorous theoretical foundations in the design and development of intervention studies. This would have an impact on the outcomes to be measured and the evaluation process. Follow-up studies with Randomized Controlled Trials (RCTs) and other research designs are required to fully comprehend the development of student career choices and preferences across STEM. The inclusion of variables beyond gender is also necessary since they provide further insights into the participation of women and other minorities in STEM pathways.

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.

Author contributions

MS and SF: review conceptualization, design, search strategy, and writing—original draft. MS, SF, MR, and B-SL: screening, study eligibility, data extraction and analysis, review, and editing. All authors contributed to the article and approved the submitted version.

Funding

The present study belongs to a research project led by the MS funded by the Spanish Ministry of Industry, Economy, and Competitiveness. Spanish State Research Agency (AEI) and European Regional Development Fund (ERDF) (Grant No. FEM2017-84589-R).

Acknowledgments

We would like to thank the UOC's library service and Efrem Melián for their support in the search of references.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.954996/full#supplementary-material

References

Acuña, S. A., Michaelis, J. E., Roth, J. D., and Towles, J. (2018). *Intervention Designed to Increase Interest in Engineering for Low-Interest, K-12 Girls did so for Boys and Girls.* Salt Lake City, UT: ASEE Annual Conference & Exposition. doi: 10.18260/1-2–30713

Arksey, H., and O'Malley, L. (2005). Scoping studies: towards a methodological framework. *Int. J. Soc. Res. Methodol. Theory Pract.* 8, 19–32. doi: 10.1080/1364557032000119616

Australian Education Council (2019). *STEM School Education Interventions: Synthesis Report.* Australian Education Council. Available online at: https://www.esa.edu.au/ (accessed July 9, 2021).

Blickenstaff, J. C. (2005). Women and science careers: leaky pipeline or gender filter? *Gend. Educ.* 17, 369–386. doi: 10.1080/09540250500145072

Brock, S. L. (2017). *Addressing the gender gap in mathematics: The Impact of Supplementary Single-Sex Mathematics Classes on Middle School Students' Mathematics Self-Efficacy, Sense of Belonging, and Achievement (Doctoral Dissertation).* Johns Hopkins University, Baltimore, MD, United States.

Brown, R. (2019). "Ethical dilemmas in education research," in *Handbook of Research Ethics and Scientific Integrity*, ed R. Iphofen, R (Cham: Springer).

Brown, R., and Brown, E. A. (2019). Estimating the effect of a teacher training program on advanced placementÂ® outcomes. *Int. J. Comput. Sci. Educ. Sch.* 2, 3–21. doi: 10.21585/ijcses.v2i4.35

Cantley, I., Prendergast, M., and Schlindwein, F. (2017). Collaborative cognitive-activation strategies as an emancipatory force in promoting girls' interest in and enjoyment of mathematics: A cross-national case study. *Int. J. Educ. Res.* 81, 38–51. doi: 10.1016/j.ijer.2016.11.004

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, 58–71. doi: 10.1007/s11199-013-0296-x

Chiu, M. S. (2011). Effects of a women-in-sciences/men-in-humanities intervention on Taiwanese adolescents' attitudes towards learning science. *Asia-Pacific Educ. Res*. 20, 322–335.

Christensen, R., and Knezek, G. (2017). Relationship of middle school student STEM interest to career intent. *J. Educ. Environ. Sci. Health*. 3, 1–13. doi: 10.21891/jeseh.275649

Christensen, R., Knezek, G., and Tyler-Wood, T. (2014). Student perceptions of science, technology, engineering and mathematics (STEM) content and careers. *Comput. Hum. Behav*. 34, 173–186. doi: 10.1016/j.chb.2014.01.046

Christensen, R., Knezek, G., and Tyler-Wood, T. (2015). Alignment of hands-on STEM engagement activities with positive STEM dispositions in secondary school students. *J. Sci. Educ. Technol*. 24, 898–909. doi: 10.1007/s10956-015-9572-6

Creswell, J. W., and Plano Clark, V. (2018). *Designing and Conducting Mixed Methods Research, 3rd edn*. Thousand Oaks, CA: Sage.

Denner, J. (2007). *The Girls Creating Games Program: An Innovative Approach to Integrating Technology Into Middle School*. Available online at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.686&rep=rep1&type=pdf

Denner, J., Werner, L., Martinez, J., and Bean, S. (2012). Computing goals, values, and expectations: Results from an after-school program for girls. *J. Women Minor. Sci. Engg*. 18, 199–213. doi: 10.1615/JWOMENMINORSCIENENG.2013002075

Drobnis, A. W. (2010). *Girls in computer science: A female only introduction class in high school (Doctoral Dissertation)*. George Mason University, Fairfax, VA, United States.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educ. Psychol*. 44, 78–89. doi: 10.1080/00461520902832368

Falco, L. D., and Summers, J. J. (2019). Improving career decision self-efficacy and STEM self-efficacy in high school girls: Evaluation of an intervention. *J. Career Develop*. 46, 62–76. doi: 10.1177/0894845317721651

Falco, L. D., Summers, J. J., and Bauman, S. (2010). Encouraging mathematics participation through improved self-efficacy: A school counseling outcomes study. *Educ. Res. Evalut*. 16, 529–549. doi: 10.1080/13803611.2011.555101

Fetters, M. D., and Molina-Azorin, J. F. (2020). Utilizing a mixed methods approach for conducting interventional evaluations. *J. Mix. Methods Res*. 14, 131–144. doi: 10.1177/1558689820912856

González-Pérez, S., Mateos de Cabo, R., and Sáinz, M. (2020). Girls in STEM: is it a female role-model thing? *Front. Psychol*. 11, 2204. doi: 10.3389/fpsyg.2020.02204

Good, C., Aronson, J., and Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *J. Appl. Dev. Psychol*. 24, 645–662. doi: 10.1016/j.appdev.2003.09.002

Grant, M. J., and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info. Libr. J*. 26, 91–108. doi: 10.1111/j.1471-1842.2009.00848.x

Hall-Lay, S. (2018). *Gender effects of robotics programs on STEM-related self-efficacy of high school students (Doctoral Dissertation)*. Walden University, Minneapolis, MN, United States.

Harackiewicz, J. M., and Priniski, S. J. (2018). Improving student outcomes in higher education: the science of targeted intervention. *Ann. Rev. Psychol*. 69, 409–435. doi: 10.1146/annurev-psych-122216-011725

Heddy, B. C. (2014). *Facilitating interest and out-of-school engagement in science in secondary school girls: Increasing the effectiveness of the teaching for transformative experience in science model through parental involvement (Doctoral Dissertation)*. University Southern California, Los Angeles, CA, United States.

Hyde, J. S., Canning, E. A., Rozek, C. S., Clarke, E., Hulleman, C. S., and Harackiewicz, J. M. (2017). The role of mothers' communication in promoting motivation for math and science course-taking in high school. *J. Res. Adolesc*. 27, 49–64. doi: 10.1111/jora.12253

Isiksal, M., and Askar, P. (2005). The effect of spreadsheet and dynamic geometry software on the achievement and self-efficacy of 7th-grade students. *Educ. Res*. 47, 333–350. doi: 10.1080/00131880500287815

Jordaan, T., and Tavenga, M. (2019). The SETH academy: enhancing learners' mathematical competence for engineering challenges in higher education. *World Trans. Eng. Technol. Educ*. 18, 284–289.

Kolne, K., and Lindsay, S. (2020). A systematic review of programs and interventions for increasing the interest and participation of children and youth with disabilities in STEM education or careers. *J. Occup. Sci*. 27, 525–546. doi: 10.1080/14427591.2019.1692692

Lazowski, R. A., and Hulleman, C. S. (2016). Motivation interventions in education: a meta-analytic review. *Rev. Educ. Res*. 86, 602–640. doi: 10.3102/0034654315617832

Leaper, C., and Brown, C. S. (2014). "Sexism in schools," in *Advances in Child Development and Behavior. The Role of Gender in Educational Context and Outcomes, Vol. 47*, eds L. S. Liben, and R. S. Bigler (London: Elsevier), 189–223. doi: 10.1016/bs.acdb.2014.04.001

Liben, L. S., and Coyle, E. F. (2014). "Developmental interventions to address the STEM gender gap," in *Advances in Child Development and Behavior. The Role of Gender in Educational Context and Outcomes*, *Vol. 47*, eds L. S. Liben, and R. S. Bigler (Amsterdam: Elsevier), 77–115. doi: 10.1016/bs.acdb.2014.06.001

Mayberry, J. (2015). *Will the integration of a girl scout STEM workshop improve the interest and confidence of undeserved millennial students in STEM? (Doctoral Dissertation)*. Trevecca Nazarene University, Nashville, TN, United States.

McHugh, L., Kelly, A. M., and Burghardt, A. (2018). Professional development for a middle school mathematics-infused science curriculum. *J. Teach. Sci. Educ*. 29, 804–828. doi: 10.1080/1046560X.2018.1514825

Mertler, C. A. (2016). *Introduction to Educational Research, 1st edn*. Thousand Oaks, CA: Sage Publications.

Mostoli, N., Rostamy, M., Shahverani, A., and Behzadi, M. H. (2019). Using the malmquist index in evaluation process to enhance mathematical literacy in high school students. *Int. J. Assess. Tool. Educ*. 6, 636–655. doi: 10.21449/ijate.623080

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., and Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol*. 18, 143. doi: 10.1186/s12874-018-0611-x

Paslov, L. S. (2006). *The effect of a piloted middle school pre-engineering program on girls' interest and achievement in mathematics (Doctoral dissertation)*. Southern Connecticut State University, New Haven, CT, United States.

Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., and McEwen, S. A. (2014). A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Res. Synth. Methods* 5, 371–385. doi: 10.1002/jrsm.1123

Prieto-Rodriguez, E., Sincock, K., and Blackmore, K. (2020). STEM initiatives matter: results from a systematic review of secondary school interventions for girls. *Int. J. Sci. Educ*. 42, 1144–1161. doi: 10.1080/09500693.2020.1749909

Rosenzweig, E. Q., and Wigfield, A. (2016). STEM motivation interventions for adolescents: a promising start, but further to go. *Educ. Psychol*. 51, 146–163. doi: 10.1080/00461520.2016.1154792

Sáinz, M. (2020). "*Brechas y Sesgos De Género En la Elección De Estudios STEM. ¿Por Qué Ocurren y Cómo Actuar Para Eliminarlas? [Gender Gaps and Biases in the Choice of STEM. Why They Occur and How to Act to Eradicate Them?]*. Sevilla: Centro de Estudios Andaluces. doi: 10.54790/actualidad.0013

Sáinz, M., Martínez-Cantos, J. L., and Meneses, J. (2020). Gendered patterns of coping responses with academic sexism in a group of Spanish secondary students. *Int. J. Soc. Psychol*. 35, 246–281. doi: 10.1080/02134748.2020.1721049

Sáinz, M., and Müller, J. (2018). Gender and family influences on Spanish students' aspirations and values in stem fields. *Int. J. Sci. Educ*. 40, 188–203. doi: 10.1080/09500693.2017.1405464

Schilling, M., and Pinnell, M. (2019). The STEM gender gap: An evaluation of the efficacy of women in engineering camps. *J. Stem. Educ. Innov. Res*. 20, 37–45. Available online at: https://www.jstem.org/jstem/index.php/JSTEM/article/view/2346/2122

Schoon, I. (2001). Teenage job aspirations and career attainment in adulthood: a 17-year follow-up study of teenagers who aspired to become scientists, health professionals, or engineers. *Int. J. Behav. Dev*. 25, 124–132. doi: 10.1080/01650250042000186

Scott, A., McAlear, F., Martin, A., and Koshy, S. (2017). Broadening participation in computing: Examining experiences of girls of color. *ACM Inroads*. 8, 48–52. doi: 10.1145/3149921

Stake, J. E. (2006). Understanding male bias against girls and women in science. *J. Appl. Soc. Psychol*. 33, 667–682. doi: 10.1111/j.1559-1816.2003.tb01918.x

Stoeger, H., Duan, X., Schirner, S., Greindl, T., and Ziegler, A. (2013). The effectiveness of a one-year online mentoring program for girls in STEM. *Comput. Educ*. 69, 408–418. doi: 10.1016/j.compedu.2013.07.032

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

Tarng, W., Lin, C. M., Liu, Y. Z., Tong, Y. N., and Pan, K. Y. (2011). A virtual reality design for learning the basic concepts of synchrotron light source. *Int. J. Comput. Eng*. 6, 175–184. doi: 10.1504/IJCSE.2011.042021

Todd, B., and Zvoch, K. (2019a). The effect of an informal science intervention on middle school girls' science affinities. *Int. J. Sci. Educ.* 41, 102–122. doi: 10.1080/09500693.2018.1534022

Todd, B., and Zvoch, K. (2019b). Exploring girls' science affinities through an informal science education program. *Res. Sci. Educ.* 49, 1647–1676. doi: 10.1007/s11165-017-9670-y

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann. Intern. Med.* 169, 467–473. doi: 10.7326/M18-0850

UNESCO (2018). *Cracking the Code. Girls' and Women's Education in Science, Technology, Engineering and Mathematics (STEM).* Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000253479 (accessed October 18, 2020).

UNESCO (2021). *A New Generation of 25 Efforts for Gender Equality in Education*. Retrieved from https://en.unesco.org/gem-report/2020genderreport (accessed December 23, 2021).

van den Hurk, A., Meelissen, M., and van Langen, A. (2019). Interventions in education to prevent STEM pipeline leakage. *Int. J. Sci. Educ.* 41, 150–164. doi: 10.1080/09500693.2018.1540897

Victor, H. V. (2005). "On the origins of expectancy theory," in *Great Minds in Management*, eds K. G. Smith and M. A. Hitt (New York, NY: Oxford University Press). pp. 239–258.

Wang, M. T., and Degol, J. (2013). Motivational pathways to STEM career choices: using expectancy–value perspective to understand individual and gender differences in STEM fields. *Dev. Rev.* 33, 304–340. doi: 10.1016/j.dr.2013.08.001

Werner, A. (2017). *How female-oriented teaching strategies impact the attitude of middle school female science students (Doctoral Dissertation)*. Frostburg State University, Frostburg, MD, United States.

Wilson, M. M. (2019). *Cultivating the roots of STEM: Investigating the influence of a STEM program on teachers' efficacy and adolescents' attitudes toward STEM (Doctoral Dissertation)*. University of Alabama, Tuscaloosa, AL, United States.

Yeager, D. S., and Walton, G. M. (2011). Social-psychological interventions in education: they're not magic. *Rev. Educ. Res.* 81, 267–301. doi: 10.3102/0034654311405999

Ziegler, A., and Heller, K. A. (2000). Effects of an attribution retraining with female students gifted in physics. *J. Educ. Gift.* 23, 217–243. doi: 10.4219/jeg-2000-585

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/fpsyg.2022.955530

OPEN ACCESS

EDITED BY
Milagros Sainz,
Open University of Catalonia, Spain

REVIEWED BY
Alicia García-Holgado,
University of Salamanca,
Spain
Andrea Vera-Gajardo,
Universidad de Valparaíso,
Chile

*CORRESPONDENCE
Clara Sainz-de-Baranda Andujar
cbaranda@hum.uc3m.es
Laura Gutiérrez-Martín
lagutier@ing.uc3m.es

SPECIALTY SECTION
This article was submitted to Gender, Sex and Sexualities, a section of the journal Frontiers in Psychology

RECEIVED 28 May 2022
ACCEPTED 09 August 2022
PUBLISHED 20 October 2022

CITATION
Sainz-de-Baranda Andujar C, Gutiérrez-Martín L, Miranda-Calero JÁ, Blanco-Ruiz M and López-Ongil C (2022) Gender biases in the training methods of affective computing: Redesign and validation of the Self-Assessment Manikin in measuring emotions *via* audiovisual clips. *Front. Psychol.* 13:955530.
doi: 10.3389/fpsyg.2022.955530

Gender biases in the training methods of affective computing: Redesign and validation of the Self-Assessment Manikin in measuring emotions *via* audiovisual clips

Clara Sainz-de-Baranda Andujar[1,2]*, Laura Gutiérrez-Martín[3]*, José Ángel Miranda-Calero[3], Marian Blanco-Ruiz[2,4] and Celia López-Ongil[2,3]

[1]Department of Communication and Media Studies, Universidad Carlos III de Madrid, Getafe, Madrid, Spain, [2]Institute on Gender Studies, Universidad Carlos III de Madrid, Getafe, Madrid, Spain, [3]Department of Electronic Technology, Universidad Carlos III de Madrid, Leganés, Spain, [4]Department of Audiovisual Communication and Advertising, Universidad Rey Juan Carlos, Fuenlabrada, Spain

Audiovisual communication is greatly contributing to the emerging research field of affective computing. The use of audiovisual stimuli within immersive virtual reality environments is providing very intense emotional reactions, which provoke spontaneous physical and physiological changes that can be assimilated into real responses. In order to ensure high-quality recognition, the artificial intelligence (AI) system must be trained with adequate data sets, including not only those gathered by smart sensors but also the tags related to the elicited emotion. Currently, there are very few techniques available for the labeling of emotions. Among them, the Self-Assessment Manikin (SAM) devised by Lang is one of the most popular. This study shows experimentally that the graphic proposal for the original SAM labelling system, as devised by Lang, is not neutral to gender and contains gender biases in its design and representation. Therefore, a new graphic design has been proposed and tested according to the guidelines of expert judges. The results of the experiment show an overall improvement in the labeling of emotions in the pleasure–arousal–dominance (PAD) affective space, particularly, for women. This research proves the relevance of applying the gender perspective in the validation of tools used throughout the years.

KEYWORDS
Self-Assessment Manikin, gender, emotion, affective space, pleasure-arousal-dominance, affective computing, artificial intelligence

Introduction

The last decades have witnessed a growing interest in the multisensorial and multimodal aspects of science and technology, the integration of the measurement of emotion through the use of smart sensors being one of the emerging research lines in fields such as communication, engineering, and psychology among others. Affective computing is based on the study, analysis, and interpretation of human emotional reactions by means of artificial intelligence (AI; Picard, 1995; Picard et al., 2001), which requires the usage of complete databases where not only the measurements from different sensors are compiled rigorously but also the tags of the experimented emotions. These tags can be unconstrained or previously predefined. The predefined ones can be discrete—chosen from a finite, predefined set of emotions—or continuous, within a predefined affective space, such as the tridimensional pleasure–arousal–dominance (PAD) space (Fontaine et al., 2016), where the experimented emotion is represented *via* numerical values on a Likert scale in the dimensions of pleasure, arousal and dominance. In any case, the tags must always be gathered while the different emotions are being elicited in volunteers *via* various external stimuli.

The most used scientific databases for the study of emotions—MANHOB (Soleymani et al., 2012) and DEAP (Koelstra et al., 2012)—use the Self-Assessment Manikin (SAM) designed by Lang (1980) and Hodes et al. (1985) in the 1980s, first as a computerised, interactive graphical interface tool, although a manual version of it was later made. In fact, this non-verbal pictorial assessment technique has generally been adopted for mapping emotions in a three-dimensional space (PAD), according to the levels of pleasure (P), arousal (A), and dominance (D) every emotion draws out of the person.

The SAM technique has been consolidated throughout the years in the scientific community as a globally reliable system to classify emotions. It provides a well-defined measure with strong psychometric properties (Lang, 1980; Bradley and Lang, 1994; Leen-Feldner et al., 2008; Olatunji et al., 2009; Soares et al., 2013; Bilsky et al., 2018). For example, in their study, Zaki and Ochsner (2015) confirm that the manikins allow people to express their emotional reactions beyond linguistic barriers or discrete labels, leveraging their empathy with the figures' expressions when observing and contemplating the image or drawing.

The SAM system provides three independent scales—PAD—associated with the emotional response to external stimuli. Each scale contains five similar figures with different expressions:

- The first scale (valence/pleasure) ranges from positive sensations to negative feelings. The farthest figure on the left shows a smile, while the one farthest to the right displays a worried/sad expression.
- The second scale (arousal/excitement) measures from the highest states of excitement to calmness. The farthest figure on the left seems ready to explode, while the one on the opposite side looks calm or asleep.
- The third scale, related to dominance, corresponds to the ability to control the intensity of the emotion experimented by the subject (Verma and Tiwary, 2015); it presents a small human figure in the square, growing from left to right, where it can be seen outside of the square because of its size.

Through these images, the person can mark any figure or space between two figures with an "x" to indicate the closest emotion to the elicited one.

For the most part, SAMs have suffered variations in the sequential order of the figures in the scales of valence and arousal, being displayed from negative to positive feelings in the case of valence and from calmness to excitement in arousal (Koelstra et al., 2012; Miranda-Correa et al., 2018). This variation in the figures' sequential order must be considered for future comparisons with results from different research papers published.

The manikins have also suffered aesthetical modifications in the figures' design (Koelstra et al., 2012; Miranda-Correa et al., 2018), up to the point of proposing the use of avatars instead of manikins (Sonderegger et al., 2016). Nonetheless, these modifications have not been validated through experimental research to the best of our knowledge, nor have they considered sociocultural or gender biases.

In this context, keeping in mind that one of the main objectives of this study is the validation of aesthetic modifications of the manikins, cultural and gender biases should be taken into consideration in the same way as the contents of video clips used to cause emotional reactions in order to generate audiovisual databases—the UC3M4Safety database for Spain (Blanco-Ruiz et al., 2021a,b) or Emotional Film for Asian culture (Deng et al., 2017). Gender and cultural differences have also been confirmed (Gantiva et al., 2011; Moltó et al., 2013) in the International Affective Picture System (IAPS; Lang et al., 2008), which includes over 1,000 pictures that represent a set of normative emotional stimuli for experimental research about attention and emotions.

The identification with human-like figures is a key concept in understanding and explaining the processes and effects that the stimuli provoke in the subjects while the experiments are being conducted. Through the figures, many emotions felt during direct encounters in personal experiences are recalled, activating what is known as autobiographical memory (Cohen, 2001; Sainz-de-Baranda et al., 2021b).

The different experiments in emotion recognition have detected that, in addition to individual differences in empathising with others (Lockwood et al., 2017; Israelashvili et al., 2019; Blanco-Ruiz et al., 2020; Sainz-de-Baranda et al., 2021a), there are also cultural, linguistic, sexual and age differences (Hagemann et al., 1999; Trommsdorff et al., 2016; Di Girolamo et al., 2019; Ge et al., 2019; Grégoire and Greening, 2020) that should be addressed and adapted so that every subject can reach a greater empathy with the audiovisual speeches being studied. In this sense, recent studies from feminist technoscience studies have highlighted that digital technologies and AI have biases in terms of gender, sex, job, class, ethnicity, and (dis)ability among others (Sumartojo et al., 2016; Hicks, 2017; Dunbar-Hester, 2019; Thaler, 2022).

Gender[1] analysis of the world around us, and thus of technology, shows that from its design to its operation, it is not gender neutral (Haraway, 1988; Harding, 1991; Wajcman, 2006; Zafra, 2011). Examples, such as the design of autonomous cars with a gender perspective to correct inequalities in the design of the traditional belt (Saleh et al., 2022), differences in cardiovascular rehabilitation (Kentner and Grace, 2017) or the John–Jennifer effect (Moss-Racussin et al., 2012), are evidence of the need for this shift towards gender sensitivity. However, this perspective must be complemented by the intersectional perspective (Crenshaw, 1991). Recent studies on the effects of AI algorithms, such as the studies by Buolamwini and Gebru (2018), Cirillo et al. (2020), Noble (2018), and Nurock (2020) among others, point out that not only gender biases are reproduced but also those of race, class, or age.

In Europe, the European Commission (2020) has incorporated the gender perspective and the intersectional perspective into research and innovation content in the Horizon Europe framework programme, with AI being one of the key areas. Examples of this line of work include projects such as VITAPATCH in Austria, which are developing a multifunctional data patch for vital and movement monitoring in everyday environments, where its researchers are integrating knowledge on feminist technoscience into the technology design process. In the case of Spain, the EMPATÍA-CM project is working to generate automatic detection mechanisms to protect victims of gender-based violence in situations of danger, and from its beginnings, it has incorporated the gender and victim perspective into its development. As Tannenbaum et al. (2019) point out; taking a gender-sensitive view improves science and technology.

In this context, and considering that one of the main objectives of this work is the validation of aesthetic modifications of the manikins, cultural and gender biases should be taken into consideration.

Materials and methods

The initial hypothesis of this research was that the tools designed and traditionally used to measure emotions, and therefore train the intelligent systems used in affective computing, were not gender neutral. For this reason, they required a methodological revision from the gender studies perspective to produce a more equal, inclusive, and diverse science.

The aim of this study was to validate aesthetic modifications to the SAMs that serve in tagging emotions within the PAD space. This question arose when the multidisciplinary UC3M4Safety team raised the need to generate an audiovisual database—the UC3M4Safety database (Blanco-Ruiz et al., 2021a,b)—to elicit emotions through audiovisual stimuli and launch an intelligent system with the ability to determine the emotional state of a person (San-Segundo et al., 2021) known as Bindi (Miranda et al., 2021). In this sense, this work focused on analysing possible gender biases in the labelling system and thus avoiding their effects in emotion recognition. It is important to note that the labelling system conditioned the resulting intelligent system because the latter is based on supervised learning.

In this section, the different aspects of the methodology followed by this research are detailed (Ortega-Toro et al., 2008). First, the protocol, the participants, and the design of the different experiments conducted are explained and, finally, the instrument of reference is shown (Supplementary material).

Protocol

In the design of questionnaires for emotional self-labelling, we have used a stepping stone of those questionnaires that are currently used in scientific databases devoted to studying emotions and that use audiovisual stimuli of different natures to elicit emotions: FilmStim (Schaefer et al., 2010), MANHOB (Soleymani et al., 2012), DEAP (Koelstra et al., 2012), and the Emotional Film database for Asian culture (Deng et al., 2017). These are among the most used and referenced ones. All of them use the SAM tool as the emotion labelling procedure in the PAD space. It is worth noting that, despite its use in these and other publications within the field, more research on the PAD model is still needed to conceive it as a solid and proved emotional dimensional model (Bakker et al., 2014). Thus, this work claims to deepen this kind of research and deals specifically with the gender bias problem within this field. To this end, the protocol followed is based on the three following phases (Figure 1):

- The first phase was aimed at acquiring the validity of the content and the form of the survey (Table 1). To this end, the

1 Gender refers to the socially defined roles, characteristics and opportunities that are considered appropriate for men, women, boys, girls and people with non-binary identities. Gender is also a product of the relationships between people and can reflect the distribution of power between them (ONU Mujeres, 2021). According to Díaz Martínez, 2016, gender perspective implies that sex and gender are reflected in research content. Gender as a perspective can manifest itself in research questions, theories, approaches, methods and dissemination. This means that sex, gender and their interaction must be adequately represented and addressed in the groups under study, and should be kept in mind if the impact of the research and the results are different. In relation to this, it is interesting to note the work developed by Anne Fausto-Sterling and Londa Schiebinger.

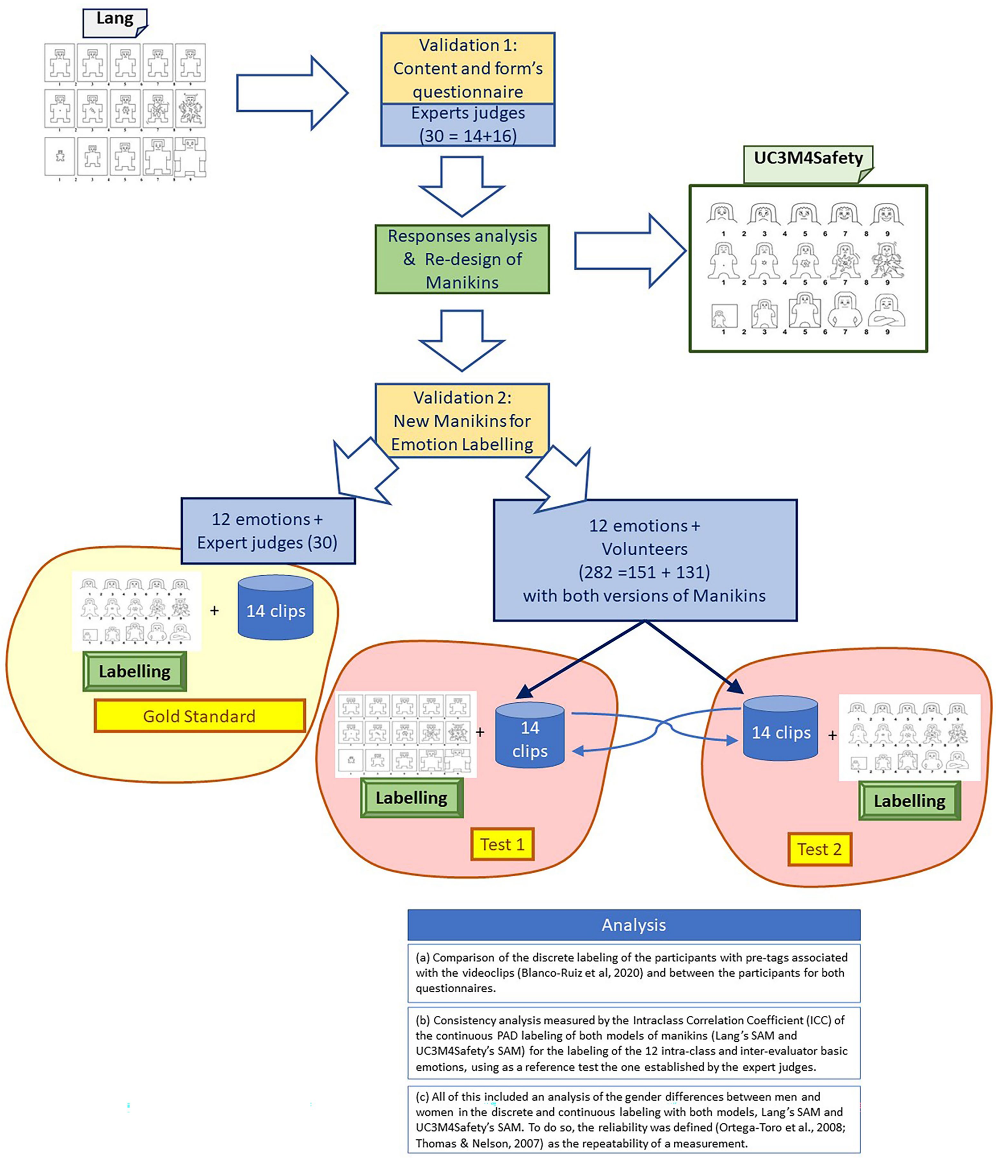

FIGURE 1
Stages and procedures involved in methodology.

questionnaire that included the SAMs with the original aesthetic designed by Lang (1980) was sent to a group of expert judges (16 women and 14 men).

- The second phase consisted of the interpretation of each of the expert judges' answers, after which the original aesthetic of the manikins was redesigned (Table 2).
- In the third phase, a two-step experiment was designed to confirm or discard the improvement in labelling between Lang's SAMs and those designed by the UC3M4Safety team (UC3M4Safety's SAMs), namely:

1. Asking the expert judges to label 12 basic emotions—described in the "Instrument" section, Table 3. This labeling

TABLE 1 Quantitative assessment issued by the expert judges about the Self-Assessment Manikins (Lang's vs. UC3M4Safety's).

Items	Lang's SAMs			UC3M4Safety's SAMs		
	Content	Form	Mean	Content	Form	Mean
Valence	0.830	0.722	0.776	0.983	0.963	0.973
Arousal	0.873	0.827	0.850	0.980	0.990	0.985
Dominance	0.867	0.643	0.755	0.993	0.980	0.987

Aiken's V coefficient for content and form validity ranges from 0 to 1.
Degree of belonging to the subject study (content). The extent to which each of the questionnaire's items was supposed to take part in the instrument was registered. To achieve this, the expert judges indicated in a scale from 0 to 10 the degree of belonging of the item to the instrument (0 = not relevant, 10 = highly relevant).
Degree of accuracy and adequacy (form). The extent to which each of the questionnaire's items accurately defined its objective was registered. Likewise, the expert judges indicated in a scale from 0 to 10, the degree of accuracy in the definition and wording of the instrument (0 = inadequate, 10 = highly adequate).

TABLE 2 Qualitative assessment issued by the expert judges about Self-Assessment Mankins of Lang (1985).

Judge	Sex	Age	Specialty	Assessment
1	Woman	52	Clinical Psychology	I think the SAMs are accurate because the body reflects the degree of arousal, and it is clearer than the face. However, I would make a change so that people can empathise better. The faces of the SAMs are very small in relation to the body, and the face should be highlighted more so that it reflects (un)happiness better and more visibly. I would remove the titles of valence, arousal and dominance.
2	Woman	57	Gender Studies/ Sociology	The images are very explicit, as a reflection from where the emotion is felt, but they are masculinised (more in the MANHOB); I would change the dummies or shapes.
3	Man	57	Clinical Psychology	The titles of valence, arousal, and dominance create confusion. Even though Lang's dummies are clear and simple, I would make them more neutral, with curves.
4	Man	45	Clinical Psychology	No comments.
5	Man	44	Psychology	No comments.
6	Woman	51	Sociology	I would remove the first "arousal" statement in the text. Arousal, nervousness and activation are easily identifiable in the manikins. I like the order from lower to higher in the shapes. Even so, I feel displeased by the drawings; I will not relate to them, especially the ones from 5 to 9.
7	Man	41	Clinical Psychology	They are fine. Consider having a male or female dummy according to the person's sex.
8	Man	45	Clinical Psychology	I am not convinced by the dummies. Perhaps a dummy should be made for men and another for women.
9	Woman	40	Clinical Psychology	There are words that may lead to an error (valence, arousal, and dominance). The drawings are good and illustrative, but a bit masculine. I would make more feminised dummies.
10	Woman	42	Clinical Psychology	The shape of the dummy is very masculine. Please improve the facial expressions. [Make] the facial expression less aggressive. Eliminate the word "dominance" and replace it with another; "dominant" refers to the dominance of a third person.
11	Woman	42	Clinical Psychology	The term "dominance" should be changed.
12	Woman	51	Communication	It is not clear when the arousal and valence categories are used; I would eliminate them. The same logic applies to dominance; I would replace it with "control." I would change the dummies so that they are more neutral.
13	Woman	31	Psychology	I would modify the manikins and highlight the faces more. Additionally, they are a bit masculine, especially when they are shown to women. I prefer the low-to-high sequential order.
14	Man	51	Communication	Consider having a male or female dummy according to the person's sex.
15	Woman	38	Psychology/ Neuroscience	Seek less robotic and masculine facial expressions. They are not relatable.
16	Woman	42	Psychology/ Neuroscience	I would prefer a more neutral set of dummies. Eliminate "arousal," "valence," and "dominance" because they are misleading. Substitute "dominion" with "control."
17	Woman	50	Gender Studies/ Sociology	Arousal has a sexual connotation; it would be better to change that word. Consider having a male or female dummy according to the sex of the participant. I would remove the title "valence" and leave "How do you feel?"
18	Man	57	Publicity	Seek an alternative to the manikins' faces, something more neutral or feminine.
19	Man	45	Clinical Psychology	It is not clear. I do not like the manikins; they are not relatable.
20	Man	44	Psychology	I would pursue a more neutral dummy. Consider using emojis.
21	Woman	51	Publicity	The images are very explicit, reflecting where the emotion is felt, but the SAMs are masculinised in all the squares.
22	Man	41	Clinical Psychology	Consider having a male or female dummy according to the sex of the person, or make something more neutral.
23	Man	45	Publicity	I would change the manikins, highlighting the faces more and making them less masculine.
24	Woman	40	Publicity	No comments.
25	Man	42	Clinical Psychology	The drawings are very good, and the graphics are illustrative, but I am not convinced by the fact that they are so masculinised.
26	Man	45	Clinical Psychology	No comments.
27	Woman	51	Communication	No comments.
28	Woman	47	Communication	Perhaps the SAMs could be redesigned for men and women specifically.
29	Man	51	Publicity	Design more neutral manikins.
30	Woman	38	Psychology/ Neuroscience	Pursue less masculine facial expressions.

TABLE 3 Classification of discrete emotions in the UC3M4Safety database (Blanco-Ruiz et al., 2021a,b).

Joy (Happiness, satisfaction)	Sadness (distress, sorrow)
Surprise (amazement, amusement)	Contempt (indifference, antipathy)
Hope (trust, safety, and faith)	Fear (distrust, anguish, and anxiety)
Attraction (desire, interest)	Disgust (repugnance, aversion)
Tenderness (Gratitude, satisfaction)	Anger (annoyance, ire, irritation, fury, and rage)
Calm (tranquillity, peace)	Tedium (boredom, weariness)

Self-elaborated.

has been used as the reference test (gold standard) in order to compare them with the labels provided by the sample.

2. Conducting an experiment where a sample of persons, divided into two groups, use both models of the SAMs under comparison to label a set of audiovisual stimuli (with emotional content); each group uses the two models of the SAMs in a different order to avoid biases.

The results validate both test A (Lang) and test B (UC3M4Safety) with the gold standard.

Sample

In the three stages of the protocol, 30 expert judges—16 women and 14 men—took part in this experiment, out of which 16 were female researchers in the fields of communication, publicity, sociology, psychology, and gender studies, and the remaining 14 were male clinical psychologists and neuropsychologists. All of them had wide professional experience (over 6 years) and knowledge of gender perspective due to their profession or tuition. The age of the participants ranged between 38 and 57 years old. All participants were Spanish speakers from the Ibero-American countries. These expert judges were asked to assess the validity of the content and the form of both manikin models (SAM Lang/SAM UC3M4Safety, Figure 2), as well as to label 12 discrete emotions selected with the SAM UC3M4Safety model (as described in the "Instrument" section, Table 3). This labeling was used as a reference test in the last phase of the experiment. The sampling method was non-probabilistic, snowball sampling. The expert judges participated voluntarily. They were informed in advance of the aims of the study and the treatment of the data collected, and they had as much time as they considered necessary.

In the third phase (2), in order to obtain the information about the labelling comparison of both manikin models (Figure 2), a sample of 282 people (151 women and 131 men) was recruited *via* an intentional sampling among students and professors in advertising and marketing studies (bachelor's and master's degrees in 2020/21 and 2021/22 academic years) from universities in the region of Madrid. The sample was between 20 and 52 (32.14 ± 9.09) years old. Previously, all were informed of the study's purpose and the treatment of the data collected. Only those who voluntarily agreed to participate in the experiment were recruited.

Before the online questionnaires were disseminated (through the Google Form platform), all participants received a lesson on measuring emotions through audiovisual stimuli and the different variables included in the SAM labelling procedure (valence, arousal, and dominance).

Afterward, those who agreed to participate voluntarily completed the questionnaire. All people were Spanish-speaking or fluent in Spanish (a prerequisite for evaluating the video clips that formed part of the sample).

The survey was conducted individually *via* each person's personal electronic devices. It was distributed during the months of October 2021 to February 2022. The average response time was 30 min.

Design

As indicated in the procedure description, section "Protocol", the study of the validity of the questionnaires that included the SAMs was conducted in the first phase, taking the "validity of the content" as the degree to which a test represented adequately its mission or objective (Wiersma, 2001; Thomas and Nelson, 2007; Ortega-Toro et al., 2008).

In order to reach optimal levels of content validity in the questionnaire designed for the collection of discrete tags (discrete emotions) and continuous tags (PAD space represented by SAM), the technique of the expert judges (Pedrosa et al., 2013) was used. To that end, these judges were asked to assess different aspects of the initial information, the measurement scale, and the questionnaire items and to perform a global assessment of each (Wiersma, 2001; Ortega-Toro et al., 2008). This process was carried out in two phases: first, Lang's SAMs were assessed, and then UC3M4Safety's SAMs, following the guidelines obtained in the first phase. Regarding each item of the instrument, the judges were asked to indicate the:

1. Degree of belonging to the subject study (content). The extent to which each item of the questionnaire was supposed to take part in the instrument was registered. To achieve this, the expert judges indicated in a scale from 0 to 10 the degree of belonging of the item to the instrument (0 = not relevant, 10 = highly relevant).
2. Degree of accuracy and adequacy (form). The extent to which each of the questionnaire's items accurately defined its objective was registered. Likewise, the expert judges indicated in a scale from 0 to 10 the degree of accuracy in the definition and wording of the instrument (0 = inadequate, 10 = highly adequate).
3. Global assessment of each item.

In the third phase, as described in the "Protocol" section, the experiment was designed to measure the validity of the labelling

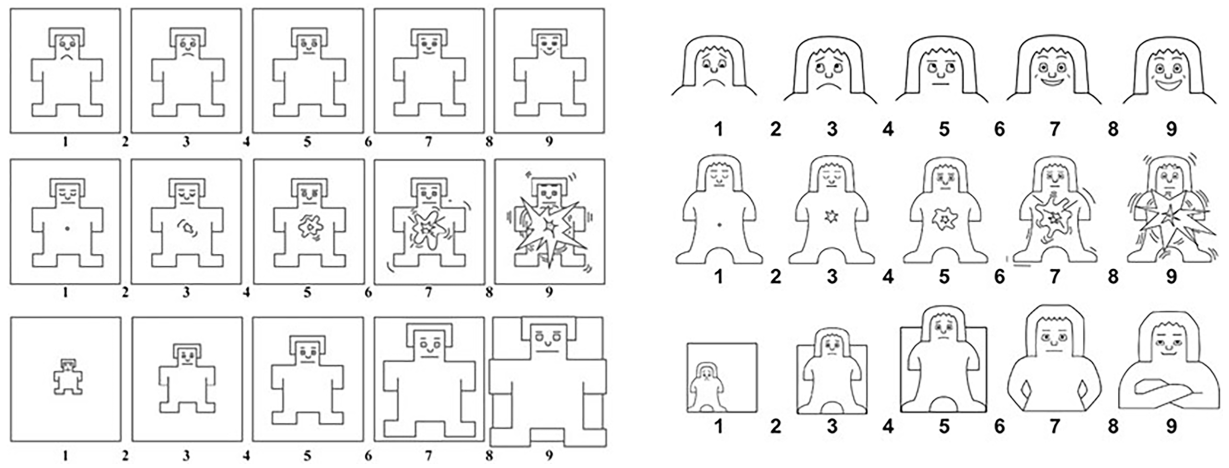

FIGURE 2
Models of the Self-Assessment Manikins proposed by Lang/UC3M4Safety.

of the new manikins (SAM UC3M4Safety) and compare them with Lang's SAM. The experiment was proposed and designed to check if the new manikins (SAM UC3M4Safety) improved the labelling procedure, leveraging the results for both genders and bringing them closer to the "golden" labels. The spirit of the experiment stemmed from the proposal by Ortega-Toro et al. (2008). The phases of the experiment were:

1. First of all, the expert judges established the references for the 12 basic emotions in the PAD tridimensional space (valence, arousal, and dominance). These basic emotions were tedium, joy, disgust, attraction, contempt, hope, tenderness, anger, fear, surprise, calm, and sadness, as described in the "Instrument" section (Table 3). Emotions were balanced between positive and negative emotions.
2. Second of all, the experiment was designed so that every participant performed two tests using Lang's SAM with a change in the sequential order as proposed by MANHOB (Soleymani et al., 2012) and DEAP (Koelstra et al., 2012) and recommended by the experts. Additionally, the UC3M4Safety SAMs were designed following the recommendations of the experts. The participants assessed each video in the three PAD dimensions, marking an "x" on each of the five figures or in any of the spaces between them, resulting in a score ranging from 1 (minimal pleasure, minimal activation, and minimal control) to 9 (maximum pleasure, maximum activation, and maximum control) per dimension.

Both questionnaires were completed by 282 participants (151 women and 131 men). The measurements were separated in time by 1 week, and they were performed in practically identical circumstances (Baumgartner, 2000).

Twelve video clips were assessed in each questionnaire, which had been previously tagged with the 12 selected basic emotions (Blanco-Ruiz et al., 2020). The videos used, one for each target emotion, were extracted from the UC3M4Safety database.[2] Two groups were created to alternate the original manikins with the new designs in order to avoid labelling biases due to the sequential order in which they were presented.

3. Finally, the responses of the participants were analysed in three aspects:

a. Comparison of the discrete labeling of the participants with pre-tags associated with the video clips (Blanco-Ruiz et al., 2020) and between the participants for both questionnaires
b. Consistency analysis measured by the intraclass correlation coefficient (ICC) of the continuous PAD labelling of both models of manikins (Lang's SAM and UC3M4Safety's SAM) for the labelling of the 12 intraclass and interevaluator basic emotions, using as a reference test the one established by the expert judges
c. All of this included an analysis of the gender differences between men and women in the discrete and continuous labeling with both models, Lang's SAM and UC3M4Safety's SAM. To do so, reliability was defined (Thomas and Nelson, 2007; Ortega-Toro et al., 2008) as the repeatability of a measurement.

Instrument

The reference instrument—a questionnaire for the labeling of the elicited emotion after viewing an audiovisual stimulus (see Supplementary Material)—was elaborated by the UC3M4Safety

2 You can access the video clips here: https://edatos.consorciomadrono.es/dataset.xhtml?persistentId=doi:10.21950/LUO1IZ

research team for the creation of an audiovisual database (Blanco-Ruiz et al., 2021a,b) and its future use to build an emotional response database capable of measuring physical (voice audio) and physiological variables (heart rate, skin temperature and conductivity, electromyogram, and breathing). The labelling questionnaire of elicited emotions *via* audiovisual stimuli consisted of a brief introduction in which the usage, the way to answer the items, the definition on the scale, and the aim of the study among others were explained. Subsequently, various sets of questions were asked about emotional response and the 12 pre-tagged audiovisual stimuli with the 12 basic emotions (Supplementary Material) were displayed to participants.

The list of emotions for this study (Table 3) was obtained from the coincidences in the Ekman studies (Ekman, 1992, 1999; Ekman and Cordaro, 2011), Izard (2016), Mauss and Robinson (2009), and Plutchik (2001), taking into account the variables used in previous audiovisual databases, such as FilmStim (Schaefer et al., 2010), MANHOB (Soleymani et al., 2012), DEAP (Koelstra et al., 2012), and Emotional Film for Asian culture (Deng et al., 2017), while incorporating the contributions from Ekman (1999, 2016) and the work of Robinson (2008) among others, in which any emotion can be represented in a positive/constructive or negative/destructive way.

Statistical analysis

The statistical analysis of data was conducted using RStudio® (RStudio, Boston, MA, United States). First, within the scope of calculating the content validity made by expert judges, Aiken's V test (Penfield and Giacobbi, 2004; Ortega-Toro et al., 2008) was used. Afterwards, in order to know the reliability of the categorical variables (discrete emotions), Kappa coefficient of Fleiss (1971) was calculated following the reference values from Altman (1991). It was an adaptation of Cohen's Kappa for evaluating the level of agreement between two or more raters. It can be expressed as follows: kappa(κ) = (Po-Pe)/(1-Pe), where Po is the observed agreement and Pe is the expected agreement.

For the continuous variables (PAD indicators), the ICC (Conroy and Metzler, 2003; Correa-Rojas, 2021) was calculated. R functions kappam.fleiss and icc from irr package were used.

Results

Expert judges: Content validity of the SAMs and PAD reference values

The quantitative assessment performed by the expert judges provided data about the validity of the content and the shape of Lang's SAM model, which signalled an Aiken's V of 0.85 in the best case (Table 1). Aiken's V values that were similar or greater than 0.8 were found both in the content of valence (0.830), arousal (0.873), and dominance (0.867). However, in terms of shape, only arousal (0.873) was higher than 0.8. Valence (0.722) and dominance (0.643) did not cross this threshold. These results showed a low assessment of the initial information.

The qualitative analysis (Table 2) provided by the expert judges contributed relevant information about the design of a new version of the SAMs: SAM UC3M4Safety.

After analysing the assessments, it was concluded that the gender biases were present in Lang's SAMs, especially in the case of dominance (the degree of control over the emotional reaction to a stimulus), alluding to the fact that the representation was very masculine, and the lines and expressions were dominant, which can be detrimental when working in emotional identification with a gender perspective.

After this result, the design of the SAMs was reviewed following the experts' guidelines, creating a seemingly more neutral model (Figure 2), and the terms used in the instructions given to the participants were also reviewed. Afterwards, the expert judges were asked once again to quantitatively assess the items that integrated the instrument, including their degree of relevance and that of precision and adequacy, as well as a global assessment of the instrument itself. The outcomes of the items related to UC3M4Safety's SAMs demonstrated a high assessment of the final information (Table 1).

In order to establish the reference values (Table 4; Figure 3) that allow the comparisons with the outcomes of the participants, the expert judges were asked to deliver the reference values for the valence, arousal, and dominance variables for each of the 12 basic emotions (Table 3) that represented the 12 basic audiovisual stimuli chosen from the UC3M4Safety audiovisual database (Blanco-Ruiz et al., 2021a,b). In Figure 3, the gold standard representation of these 12 emotions is presented in three-dimensional PAD space, which places every emotion in a low-medium-high level of excitement, pleasure, and dominance.

TABLE 4 Reference values established by the expert judges (Likert 1–9).

Emotion	Mean valence (standard deviation)	Mean arousal (standard deviation)	Mean dominance (standard deviation)
Tedium	3.00 (0.00)	1.07 (0.25)	6.23 (0.90)
Joy	8.00 (0.00)	7.00 (0.00)	6.97 (0.18)
Disgust	1.93 (0.25)	7.07 (0.25)	2.47 (0.51)
Attraction	8.00 (0.00)	7.00 (0.00)	6.53 (0.51)
Contempt	3.13 (0.35)	5.13 (0.51)	7.73 (0.69)
Hope	7.00 (0.00)	2.00 (0.00)	6.87 (0.51)
Tenderness	8.20 (0.41)	3.87 (0.51)	9.00 (0.00)
Anger	1.07 (0.25)	8.93 (0.25)	6.13 (1.17)
Fear	1.00 (0.00)	9.00 (0.00)	1.47 (0.51)
Surprise	5.93 (0.25)	7.93 (0.25)	3.00 (0.00)
Calm	6.93 (0.25)	1.00 (0.00)	9.00 (0.00)
Sadness	1.00 (0.00)	3.57 (1.28)	4.40 (1.28)

Mean of the reported values by the experts for the three different dimensions of the PAD space and their standard deviation (between brackets)

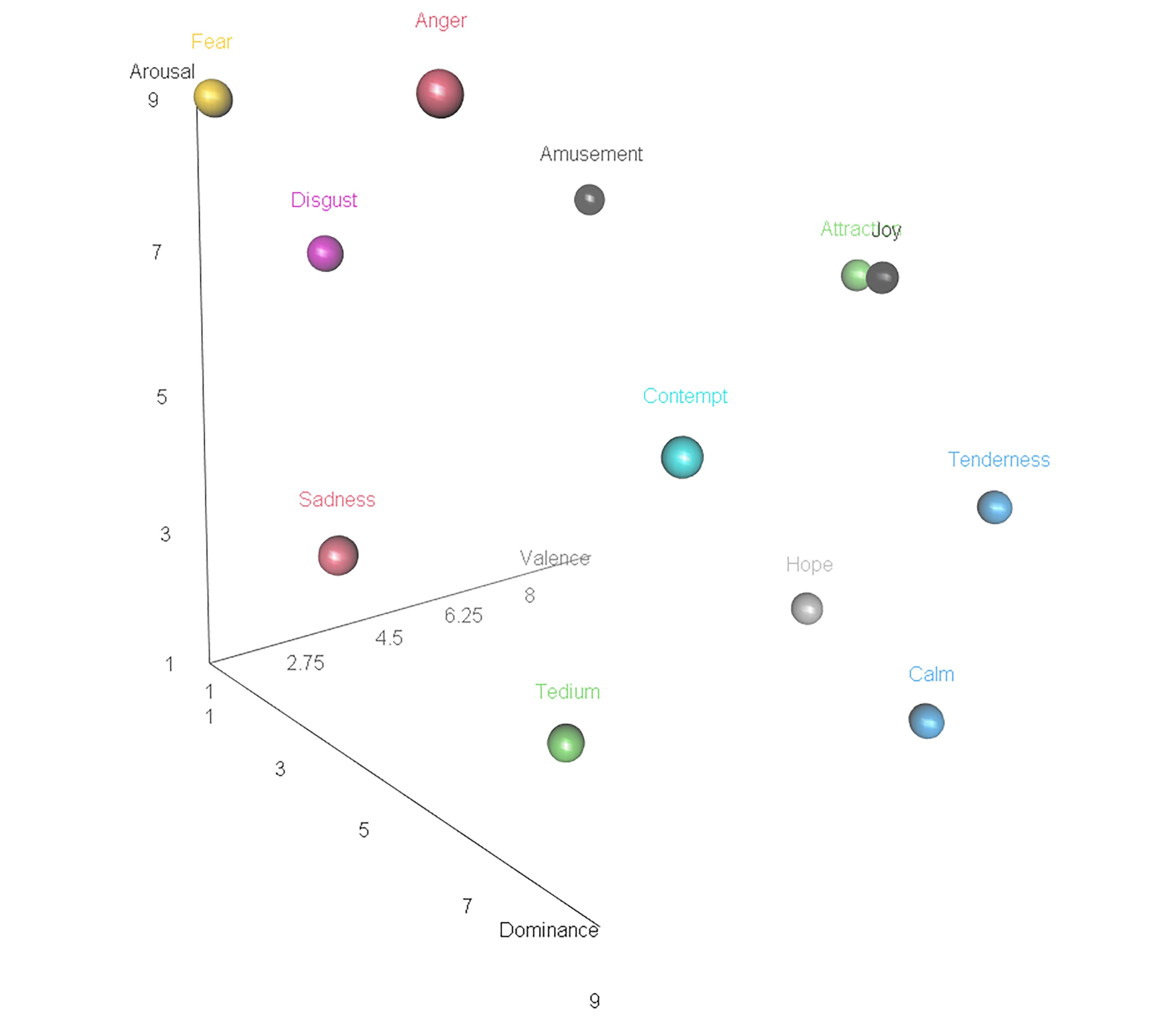

FIGURE 3
Representation in the pleasure–arousal–dominance space of the reference values established by the expert judges (gold standard). The colours are just to help to identify which point represents each emotion. This representation presents the gold standard in the three-dimensional pleasure–arousal–dominance (PAD) space and places each emotion in a low-medium-high level of excitement, pleasure, and dominance.

Experiment results

Validity and consistency of the discrete-labeling emotions

With the intent of confirming the agreement between the 12 emotions under study (Table 3) that represented the 12 previously tagged audiovisual stimuli (Blanco-Ruiz et al., 2021a,b) and those reported by the participants, a study was conducted using Kappa coefficient of Fleiss (1971). This coefficient measured the degree of correlation among raters of the nominal categories when the same samples were evaluated. The global results showed indices between 0.841 and 0.97 (Table 5) with practically no variation (delta). These results confirmed that the audiovisual stimuli, independent of the assessment system of manikins, generated an emotion in a unique fashion.

From a gender perspective, we observed that men obtained results with almost no variation (delta) and sustained Kappa index values between 0.97 and 1, that is, they showed practically perfect agreement. Women obtained a Kappa index higher than 0.7, which is a good level of agreement. However, this result confirmed that women have greater variability than men. An improvement was observed in the discrete labelling for women and, to a lesser extent, for men as well when the UC3M4Safety SAMs were used in the questionnaires to classify the experienced emotions.

Validity and consistency of emotions of the continuous labeling (pleasure–arousal–dominance)

Once the existence of a high level of agreement between the participants when labelling using discrete emotions was confirmed, the consistency of the continuous tags used for every emotion by the participants was analysed. This analysis considered intraclass and interassessor consistency, that is, if there was a variation in the measurements made by the instrument about the same topic in the same conditions. For this purpose, the ICC was used with the single-rating, absolute-agreement, Two-Way Mixed Effects Model (Table 6). The results corroborated the changes that were taking place in the continuous labelling (PAD) from Lang's model to UC3M4Safety's model.

Afterwards, for every emotion provided by the expert judges, agreement with the reference test (golden test) was evaluated (Table 7) in an independent manner for every participant (Figure 4), utilising the ICC index with the single-rating, absolute-agreement, Two-Way Random-Effects Model for each of the labelling methods. The results showed an increase in consistency and agreement between the data corresponding to UC3M4Safety's SAMs, increasing the ICC to 0.21, 0.22, or 0.23 in the emotions of joy, attraction and surprise, respectively. Additionally, due to that greater agreement, it could be observed that the position of the emotions in the PAD space was more closely adjusted to the one reported by the expert judges, and had a lower standard deviation.

Finally, the greater agreement found for UC3M4Safety's SAMs was studied. In order to do this, the data reported with UC3M4Safety's SAMs and Lang's SAMs were analysed, comparing them to the golden labels provided by the expert judges in an individual way for every participant.

Women started off with worse data with Lang's SAMs to obtaining better results than men according to UC3M4Safety's SAMs. In Figure 4, the mean correlation index of the 12 emotions for each of the participants in relation to the reference test for both models, as can be observed in almost all cases as a dotted yellow line, is above the blue one, meaning the agreement between the gold standard set by the experts and the participants is higher using the new methodology. Moreover, these results show that there was a greater consistency in the data in relation to the reference (golden) test when the UC3M4Safety SAMs were used, especially in the case of women. Out of 57 participants that obtained the same ICC results with both manikins, only six were women.

Discussion

This research started from the hypothesis that the tools traditionally used to measure emotions, and therefore train intelligent systems used in affective computing, were not gender neutral. In particular, whether the SAM instrument as a methodology could be considered a neutral tool was evaluated.

The results have shown that the manikins (SAMs), despite being designed with the objective of being neutral, are not perceived as such by the participants. In particular, the case of the graphic representation of dominance is paradigmatic since what is understood as neutral is perceived as a masculine trait. This particular result is not isolated but is part of a mainstream in scientific knowledge and technology that takes the androcentric point of view as neutral (Leavy, 2018). As Haslanger (2000) points out, in science and innovation, men are the norm and women are deviations from it.

The United Nations Organisations (ONU Mujeres, 2021, para. 3) define gender perspective as 'the assessment process of the consequences for women and men of any planned activity, including laws, policies or programs, in all sectors and at all levels'. The European Commission—the Directorate-General for Research and Innovation—and currently the State Research Agency (Agencia Estatal de Investigación) in Spain argue that engaging the gender research dimension 'implies that gender is considered a key analytical and explanatory variable in research' (Dirección General de Investigación e Innovación, 2011, p. 10). This study corroborates the importance of applying the gender perspective so that results are not partial and constitute quality, egalitarian research.

Technology development is increasingly influencing the behaviour of people in everyday life. However, according to Leavy (2018) and Wajcman (2006), the over-representation of men in the design of these technologies could perpetuate gender inequality. Different researchers have demonstrated that AI algorithms are not neutral and contribute to reproducing existing biases in today's society, the most evident being those of gender and race (O'Neil, 2016; Buolamwini and Gebru, 2018; Noble, 2018; Cirillo et al., 2020). The main types of biases in AI include gender, ethnicity, and age, and these can increase social inequalities or discrimination. Furthermore, these biases affect all sectors in which AI intervenes—from resource allocation in healthcare, justice, education, or employment—and concern both sectors that may look anecdotal—and are not in any way—and relational machines (especially with personal assistants) or vehicles with integrated voice recognition systems (Nurock, 2020).

A clear example is the controversial area of the application of AI in facial recognition software used by law enforcement agencies (Domingo, 2021). Buolamwini and Gebru (2018) proved that the software utilised by the police in the United States had an error rate regarding gender, ethnicity, and age. This error rate clearly favoured young, white men, while negatively affecting black, elderly women.

The newest line in the measurement of emotions for the prediction of scenarios and human behaviour allows interdisciplinary work between disciplines, such as social sciences and engineering, with the aim of making new technologies increasingly "more human." The applicability of this interdisciplinary synergy that is being applied intends to improve scientific knowledge by introducing the gender perspective into the design of technologies and into the selection of data to train algorithms (Sainz-de-Baranda et al., 2021a, 2022).

The incorporation of areas such as communication with gender perspective in the processes of research of technology and AI allows the advancement of technological development towards solutions that really improve people's lives (Rituerto-González et al., 2019, 2020; Sainz-de-Baranda et al., 2021a, 2022; Miranda et al., 2022).

Audiovisual communication is greatly contributing to the emerging research field of affective computing. Within

TABLE 5 Fleiss' Kappa index for the measurement of consistency of experienced discrete emotions with both Self-Assessment Manikin models.

Emotion	Global			Women			Men		
	Lang SAM	UC3M4Safety SAM	Delta	Lang SAM	UC3M4Safety SAM	Delta	Lang SAM	UC3M4Safety SAM	Delta
Tedium	0.841	0.828	−0.013	0.739	0.724	−0.015	0.983	0.983	0.000
Joy	0.911	0.911	0.000	0.845	0.845	0.000	0.992	0.992	0.000
Disgust	0.892	0.886	−0.006	0.820	0.811	−0.009	0.992	0.992	0.000
Attraction	0.870	0.893	0.023	0.768	0.809	0.041	0.992	0.992	0.000
Contempt	0.878	0.909	0.031	0.784	0.832	0.048	0.992	1.000	0.008
Hope	0.939	0.919	−0.020	0.893	0.858	−0.035	0.992	0.992	0.000
Tenderness	0.924	0.924	0.000	0.868	0.869	0.001	0.992	0.992	0.000
Anger	0.908	0.916	0.008	0.831	0.844	0.013	1.000	1.000	0.000
Fear	0.945	0.945	0.000	0.903	0.903	0.000	1.000	1.000	0.000
Surprise	0.860	0.875	0.015	0.744	0.769	0.025	1.000	1.000	0.000
Calm	0.872	0.900	0.028	0.793	0.839	0.046	0.976	0.976	0.000
Sadness	0.970	0.966	−0.004	0.951	0.938	−0.013	0.992	1.000	0.008

Fleiss' Kappa coefficient ranges from − 1 to + 1. Negative values represent that the agreement is lower than the expected by chance. On the other hand, positive values imply the rater agreement exceeds chance agreement. Within the positive range, values above 0.80, values between 0.40 and 0.80, and values below 0.40 represent excellent, fair and poor agreement, respectively.

TABLE 6 Assessment of the intraclass pleasure–arousal–dominance for each emotion with both Self-Assessment-Manikin models.

Emotion	Global			Women			Men		
	ICC Lang SAM	ICC UC3M4Safety SAM	Delta	ICC Lang SAM	ICC UC3M4Safety SAM	Delta	ICC Lang SAM	ICC UC3M4Safety SAM	Delta
Tedium	0.8675	0.9628	0.095	0.8531	0.9897	0.137	0.8891	0.9359	0.047
Joy	0.5790	0.6700	0.091	0.7803	0.7513	−0.029	0.3626	0.6233	0.261
Disgust	0.8081	0.9356	0.127	0.7455	0.9242	0.179	0.8971	0.9612	0.064
Attraction	0.3188	0.8195	0.501	0.5889	0.9838	0.395	0.5336	0.8349	0.301
Contempt	0.8721	0.9701	0.098	0.7887	0.9952	0.206	0.9831	0.9840	0.001
Hope	0.8066	0.9752	0.169	0.7922	0.9977	0.205	0.8396	0.9531	0.113
Tenderness	0.8246	1.0000	0.175	0.7221	1.0000	0.278	0.9561	1.0000	0.044
Anger	0.9126	0.9685	0.056	0.9212	0.9758	0.055	0.9575	0.9880	0.03
Fear	0.9380	0.9896	0.052	0.8940	0.9936	0.100	0.9882	0.9932	0.005
Surprise	0.7603	0.9561	0.196	0.8043	0.9867	0.182	0.7199	0.9254	0.205
Calm	0.9507	0.9899	0.039	0.9187	0.9914	0.073	0.9862	0.9884	0.002
Sadness	0.6372	0.8526	0.215	0.5527	0.9850	0.432	0.7429	0.9225	0.180
Mean	0.7729	0.9242	0.151	0.7802	0.9645	0.184	0.8213	0.9258	0.104

ICC with the single-rating, absolute-agreement, Two-Way Mixed Effects Model. Values range from 0 to 1. Below 0.50, between 0.50 and 0.75, between 0.75 and 0.90, and above 0.90, the correlation is considered poor, moderate, good, and excellent, respectively. Delta variable is the difference between ICC UC3M4Safety SAM and ICC Lang SAM. The results are disaggregated by gender.

immersive virtual reality environments, the elicitation of emotions *via* audiovisual stimuli is showing very intense emotional reactions that can be assimilated into real ones in terms of physical and physiological bio-signals (Blanco-Ruiz et al., 2020; Miranda et al., 2021). However, in order to guarantee a high-quality emotional recognition, the AI system must be trained with adequate data sets, including not only those collected by smart sensors but also the tags related to the elicited emotion. Currently, there are very few techniques available to label emotions. Among them, the SAM, which was created by Lang (1980) and Hodes et al. (1985), is one of the most popular.

The results of this study show that the fact that gender socialisation grants differentiating roles to men and women is not considered. These roles start in childhood, from their initiation in social and cultural life, and are reinforced by the influence of socialising agents. Certain cognitive, attitudinal, and behavioural styles are adopted as well as axiological codes and stereotypical morals and rules that follow the social conduct assigned to each gender (Bosch and Ferrer-Pérez, 2002). The trend of identifying

TABLE 7 Degree of agreement between the continuous labelling comparison of the participants with the gold standard for each of the emotions.

Emotion	ICC		Model	Mean valence (standard deviation)	Mean arousal (standard deviation)	Mean dominance (standard deviation)
	Lang SAM	UC3M4Safety SAM				
Tedium	0.912	0.979	Ref.	3.00 (0.00)	1.07 (0.25)	6.23 (0.90)
			Lang SAM	3.62 (1.00)	1.31 (1.01)	6.88 (1.47)
			UC3M4Safety SAM	3.05 (0.32)	1.01 (0.10)	6.26 (0.85)
Joy	0.650	0.861	Ref.	8.00 (0.00)	7.00 (0.00)	6.97 (0.18)
			Lang SAM	8.28 (0.52)	7.31 (1.03)	6.01 (1.46)
			UC3M4Safety SAM	7.99 (0.56)	7.03 (0.38)	6.71 (0.76)
Disgust	0.869	0.958	Ref.	1.93 (0.25)	7.07 (0.25)	2.47 (0.51)
			Lang SAM	2.48 (1.19)	7.02 (1.01)	3.26 (1.81)
			UC3M4Safety SAM	2.15 (0.66)	7.06 (0.54)	2.74 (1.11)
Attraction	0.656	0.873	Ref.	8.00 (0.00)	7.00 (0.00)	6.53 (0.51)
			Lang SAM	7.50 (0.83)	6.82 (0.57)	6.41 (1.01)
			UC3M4Safety SAM	7.97 (0.18)	6.98 (0.14)	6.53 (0.57)
Contempt	0.929	0.984	Ref.	3.13 (0.35)	5.13 (0.51)	7.73 (0.69)
			Lang SAM	3.46 (0.83)	4.56 (1.20)	7.93 (0.78)
			UC3M4Safety SAM	3.05 (0.27)	5.01 (0.12)	7.58 (0.63)
Hope	0.860	0.981	Ref.	7.00 (0.00)	2.00 (0.00)	6.87 (0.51)
			Lang SAM	7.51 (0.82)	2.88 (1.65)	6.83 (1.05)
			UC3M4Safety SAM	7.06 (0.24)	2.14 (0.68)	6.99 (0.29)
Tenderness	0.894	0.999	Ref.	8.20 (0.41)	3.87 (0.51)	9.00 (0.00)
			Lang SAM	8.14 (0.59)	3.64 (1.22)	8.28 (1.65)
			UC3M4Safety SAM	8.00 (0.00)	4.00 (0.00)	9.00 (0.00)
Anger	0.946	0.982	Ref.	1.07 (0.25)	8.93 (0.25)	6.13 (1.17)
			Lang SAM	1.34 (0.70)	8.49 (0.90)	6.05 (1.62)
			UC3M4Safety SAM	1.11 (0.32)	8.83 (0.52)	6.31 (1.09)
Anger	0.947	0.992	Ref.	1.00 (0.00)	9.00 (0.00)	1.47 (0.51)
			Lang SAM	1.29 (0.63)	8.54 (0.96)	2.14 (1.44)
			UC3M4Safety SAM	1.09 (0.31)	8.93 (0.32)	1.50 (0.67)
Surprise	0.726	0.960	Ref.	5.93 (0.25)	7.93 (0.25)	3.00 (0.00)
			Lang SAM	6.21 (0.80)	7.77 (1.12)	3.56 (1.73)
			UC3M4Safety SAM	5.96 (0.52)	7.93 (0.47)	3.10 (0.63)
Calm	0.965	0.994	Ref.	6.93 (0.25)	1.00 (0.00)	9.00 (0.00)
			Lang SAM	6.72 (1.23)	1.30 (0.85)	8.66 (0.93)
			UC3M4Safety SAM	7.05 (0.59)	1.06 (0.37)	8.95 (0.31)
Sadness	0.830	0.893	Ref.	1.00 (0.00)	3.57 (1.28)	4.40 (1.28)
			Lang SAM	1.16 (0.66)	3.90 (1.65)	4.09 (1.20)
			UC3M4Safety SAM	1.01 (0.08)	3.05 (0.35)	3.80 (0.98)

ICC with the single-rating, absolute-agreement, and Two-Way Random-Effects Model for each of the labelling methods. Values range from 0 to 1. Below 0.50, between 0.50 and 0.75, between 0.75 and 0.90, and above 0.90, the correlation is considered poor, moderate, good, and excellent, respectively. Mean values for the three different dimensions of the PAD space and their standard deviation (between brackets). Reference (ref.) model corresponds to the values reported by the experts. Lang SAM refers to the values reported using Lang's manikin's questionnaire and UC3M4Safety's SAM as the redesigned one.

people with their peers—or those just like them—(Igartua and Muñiz, 2008; Soto-Sanfiel et al., 2010) has added to the learning of emotions according to individual experiences, which can serve as an explanation for the existing discrepancy in the discrete labelling between men and women. Men have obtained more favourable results, with a high level of agreement, while women have greater variability. Even though discrete tags are not variable and generally have a high level of agreement with previously reported ones, a raise in the level of agreement when questionnaires containing UC3M4Safety's SAMs are used has been observed, thus clarifying the new design of manikins when participants experience an emotion during the watching/visualisation of a video after assessing the rest of the PAD characteristics of emotion – especially for women.

In the case of the analysis of emotions reported in a numerical way by the participants and which were represented in a

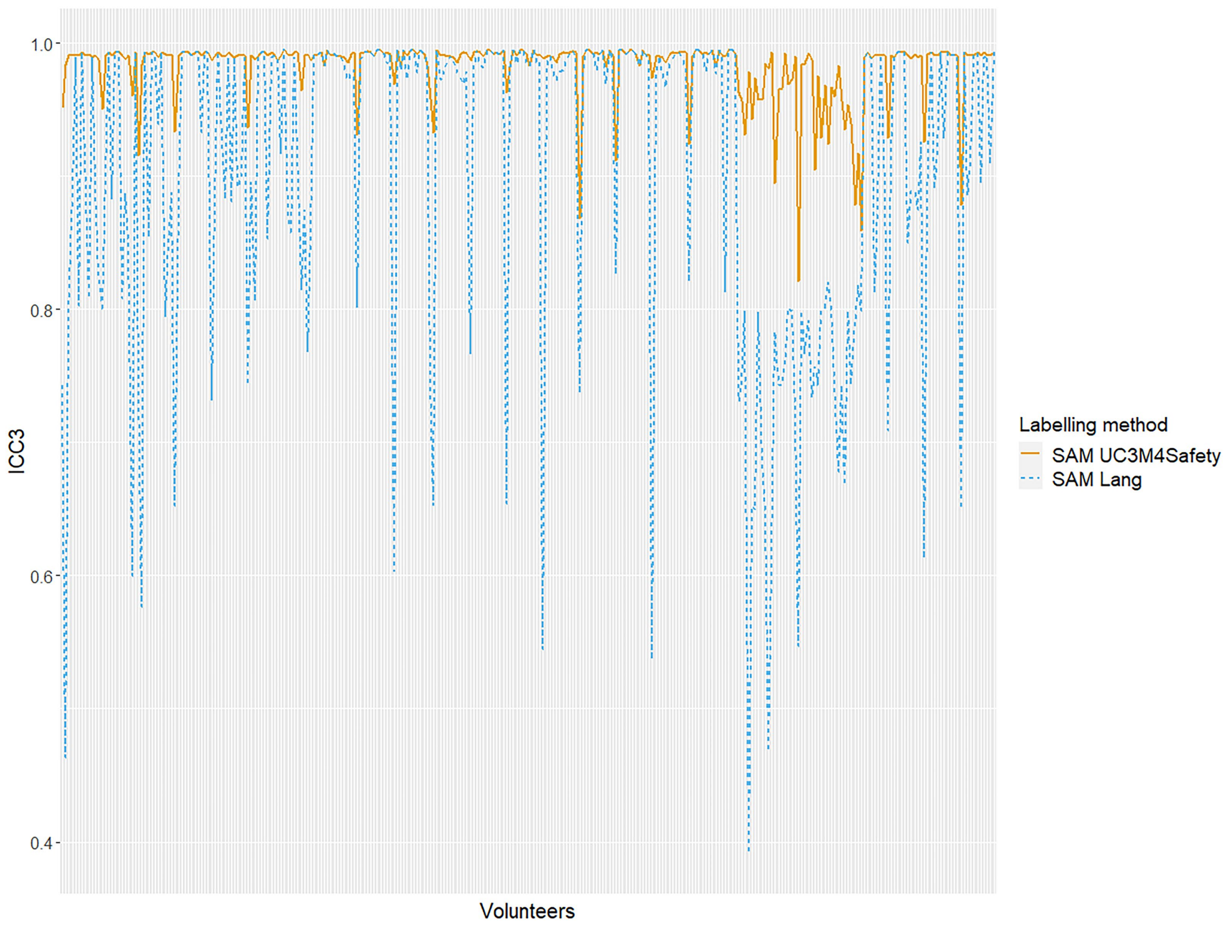

FIGURE 4
Mean intraclass correlation index of the 12 emotions for each of the participants in relation to the reference test for both models. The *y*-axis represents the mean intraclass correlation coefficient (ICC) value for the 12 emotions with respect to the gold standard. The *x*-axis represents each of the volunteers by identifier. The yellow line shows the results corresponding to answers collected using the UC3M4Safety SAM labelling questionnaire. On the other hand, the blue dotted line presents the values obtained by means of the Lang SAM questionnaire.

tridimensional fashion in the PAD affective space (valence, activation, and dominance), the differences between the tagged emotion *a priori* and those reported by gender were bigger if both SAM models were applied.

The labelling process of each emotion in the PAD space using the UC3M4Safety SAMs had a higher degree of coincidence with the reference test (gold standard) than that of Lang's SAMs, both in men and women. These results prove the UC3M4Safety SAM as a reliable and useful tool for the assessment of emotions.

An intersectional feminist approach to new technologies exposes the discriminatory biases of gender, race, and class in the generation and usage of data through information communication technologies (D'Ignazio and Klein, 2020; Blanco-Ruiz, 2022). These results make the inclusion of the gender perspective an imperative in the design of technology and in the generation of databases that are used to train AI systems that coincide with the proposal made by Revi Sterling (2013), who criticises the fact that women, as potential beneficiaries of those technologies, continue to be excluded in design processes.

As pointed out by Schiebinger (2021), identifying gender bias and understanding how it operates is crucially important, "but analysis cannot stop there" (p.3). Future technological developments should be influenced by an intersectional feminist approach (Crenshaw, 1991) in order to avoid reproducing discriminatory gender, race, and class biases, not only in design but also in use (D'Ignazio and Klein, 2020; Blanco-Ruiz, 2022). Incorporating sex, gender, and intersectionality analysis in research is a crucial component that contributes to science and technology (Tannenbaum et al., 2019). Companies such as Google, Amazon, and Facebook are beginning to be aware of the benefits of these inclusive policies. Still, the change must go further; it must permeate the three domains of scientific infrastructure: funding agencies, peer-reviewed journals, and universities (Schiebinger, 2021).

This study is also limited by its own cultural context; it should be tested in other countries to see if the gendered re-reading of the SAM that has been carried out in this study also works in other cultural contexts.

Conclusion

The new version of UC3M4Safety's SAMs considers gender perspective in its design and its contribution to the communication field, which allows for the generation of databases that enable better creation of AI systems (affective computing) in order to improve quality of life and avoiding gender biases for both women and men.

The need to revise the procedures used for decades in science—and more concretely, in AI—in order to avoid biases of any kind due to age, ethnicity, gender, or others is left on record.

It has been confirmed that Lang's SAMs contain gender biases and, consequently, the data resulting from the labelling of emotional reactions that former studies used based on audiovisual databases may be biased, and the generated AI systems could be identifying emotions incorrectly from the analysis of these bio-signals.

This type of research could serve as an inspiration to increase the interest of young people, especially women, in Science, Technology, Engineering, and Mathematics (STEM) fields, as it shows how a small change in the representation of a measuring instrument, such as the SAM, could mean that the perception of half of the population is not considered. Audiovisual and emotions are very attractive areas for young people and can serve as magnets to attract their attention to other possibilities of transferring knowledge to society through the STEM disciplines and their cooperation with other areas of knowledge. The national and international equality policies that foster inclusion of the gender dimension in research and that propel interdisciplinary work—which in our case is that of communication, gender studies, and engineering—produce breakthroughs to develop a more egalitarian scientific knowledge.

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://edatos.consorciomadrono.es/dataverse/empatia.

Ethics statement

The studies involving human participants were reviewed and approved by Universidad Carlos III de Madrid. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CS contributed to the study conception and design, did the experiment, performed the material preparation and data analysis, wrote the first draft of the manuscript and commented on previous versions of the manuscript, and read and approved the final manuscript. LG-M performed the material preparation and data analysis, wrote the first draft of the manuscript and commented on previous versions of the manuscript, and read and approved the final manuscript. JM-C and MB-R contributed to the study design, did the experiment, and commented on previous versions of the manuscript. CL-O contributed to the study conception and design, wrote the first draft of the manuscript and commented on previous versions of the manuscript, and read and approved the final manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Department of Research and Innovation of Madrid Regional Authority under Grant EMPATÍA-CM:Y2018/TCS-5046; and State Research Agency (Spain) under grant PID2019-106695RB-I00/AI-GENBIAS/10.13039/501100011033.

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.955530/full#supplementary-material

References

Altman, D. G. (1991). *Practical Statistics for Medical Research.* London (UK): Chapman and Hall.

Bakker, I., van der Voordt, T., Vink, P., and de Boon, J. (2014). Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Curr. Psychol.* 33, 405–421. doi: 10.1007/s12144-014-9219-4

Baumgartner, T. A. (2000). Estimating the stability reliability of a store. *Meas. Phys. Educ. Exerc. Sci.* 4, 175–178. doi: 10.1207/S15327841Mpee0403_3

Bilsky, S. A., Feldner, M. T., Bynion, T., Rojas, S. M., and Leen-Feldner, E. W. (2018). Child anxiety and parental anxiety sensitivity are related to parent sick role reinforcement. *Parenting* 18, 110–125. doi: 10.1080/15295192.2018.1444132

Blanco-Ruiz, M. (2022). "Perspectiva de género en el entorno digital [Gender perspective in the digital environment]," in *Curación Digital and Género en la Ciencia de la Información: Acceso and Preservación*. eds. M. J. Vicentini, R. San-Segundo, J. A. F. Montoya, D. Martínez-Ávila and L. A. Landim (Salamanca (Spain): Ediciones Universidad de Salamanca), 45–59.

Blanco-Ruiz, M., Gutiérrez, L., Miranda, J. A., Canabal, M. F., Romero, E., Sainz-de-Baranda, C., et al. (2021a). UC3M4Safety database—list of audiovisual stimuli.

Blanco-Ruiz, M., Gutiérrez, L., Miranda, J. A., Canabal, M. F., Romero, E., Sainz-de-Baranda, A.C., et al. (2021b). UC3M4Safety database—list of audiovisual stimuli [video]

Blanco-Ruiz, M., Sainz-de-Baranda, A. C., Gutiérrez-Martín, L., Romero-Perales, E., and López-Ongil, C. (2020). Emotion elicitation under audiovisual stimuli reception: should artificial intelligence consider the gender perspective? *Int. J. Environ. Res. Public Health* 17:8534. doi: 10.3390/ijerph17228534

Bosch, F. E., and Ferrer-Pérez, V. A. (2002). *La voz de las Invisibles: Las Víctimas de un mal Amor Que Mata [The Voice of the Invisible: Victims of a Bad Love That Kills]*. Valencia (Spain): Cátedra.

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Buolamwini, J., and Gebru, T. (2018). Gender shades: intersectional accuracy disparities in commercial gender classification. *Proc. Machine Learn. Res.* 81, 1–15.

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., et al. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digit. Med.* 3:81. doi: 10.1038/s41746-020-0288-5

Cohen, J. C. (2001). Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Commun. Soc.* 4, 245–264. doi: 10.1207/S15327825MCS0403_01

Conroy, D. E., and Metzler, J. (2003). Temporal stability of performance failure appraisal inventory items. *Meas. Phys. Educ. Exerc. Sci.* 7, 243–261. doi: 10.1207/S15327841MPEE0704_3

Correa-Rojas, J. (2021). Coeficiente de correlación intraclase: Aplicaciones Para estimar la estabilidad temporal de un instrumento de medida [Intraclass correlation coefficient: applications for estimating the temporal stability of a measuring instrument]. *Ciencias Psicol.* 15, 1–16. doi: 10.22235/cp.v15i2.2318

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Rev.* 43, 1241–1299. doi: 10.2307/1229039

D'Ignazio, C., and Klein, L. F. (2020). *Data Feminism.* Cambridge, MA (USA): MIT Press.

Deng, Y., Yang, M., and Zhou, R. (2017). A new standardized emotional film database for Asian culture. *Front. Psychol.* 8, 56–66. doi: 10.3389/fpsyg.2017.01941

Di Girolamo, M., Giromini, L., Winters, C. L., Serie, C. M. B., and de Ruiter, C. (2019). The questionnaire of cognitive and affective empathy: a comparison between paper-and-pencil versus online formats in Italian samples. *J. Pers. Assess.* 101, 159–170. doi: 10.1080/00223891.2017.1389745

Díaz Martínez, C. (2016). "La perspectiva de género en investigación social [Gender perspective in social research]," in *El análisis de la realidad social: métodos y técnicas de investigación*. ed. Alianza (Madrid (Spain)), 176–201.

Dirección General de Investigación e Innovación (2011). *Toolkit Gender in EU-Funded Research.* Brussels (Belgium): Publications Office of the European Union.

Domingo, J. C. (2021). Utilización del sistema de reconocimiento facial Para preservar la seguridad ciudadana [Use of the facial recognition system to preserve public safety]. *El Criminal. Digit.* 9, 20–37.

Dunbar-Hester, C. (2019). *Hacking Diversity: The Politics of Inclusion in Open Technology Cultures.* Princeton, New Jersey (United States): Princeton University Press.

Ekman, P. (1992). Are there basic emotions? *Psychol. Rev.* 99, 550–553. doi: 10.1037/0033-295X.99.3.550

Ekman, P. (1999). "Basic emotions," in *Handbook of Cognition and Emotion*. eds. T. Dalgleish and M. J. Power (Hoboken, New Jersey (United States): John Wiley and Sons Ltd.), 45–60.

Ekman, P. (2016). What scientists who study emotion agree about. *Perspect. Psychol. Sci.* 11, 31–34. doi: 10.1177/1745691615596992

Ekman, P., and Cordaro, D. (2011). What is meant by calling emotions basic? *Emot. Rev.* 3, 364–370. doi: 10.1177/1754073911410740

European Commission (2020). Gendered innovations 2: how inclusive analysis contributes to research and innovation: policy review. Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2777/316197

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychol. Bull.* 76, 378–382. doi: 10.1037/h0031619

Fontaine, J. R., Klaus, R. S., Etienne, B. R., and Phoebe, C. E. (2016). The world of emotions is not two-dimensional. *Psychol. Sci.* 18, 1050–1057. doi: 10.1111/j.1467-9280.2007.02024.x

Gantiva, C., Guerra, P., and Vila, J. (2011). Validación colombiana del sistema internacional de imágenes afectivas: Evidencias del origen transcultural de la emoción [Colombian validation of the international affective imagery system: evidence of the cross-cultural origin of emotion]. *Acta Colombiana de Psicología* 14, 103–111.

Ge, Y., Zhao, G., Zhang, Y., Houston, R. J., and Song, J. (2019). A standardised database of Chinese emotional film clips. *Cognit. Emot.* 33, 976–990. doi: 10.1080/02699931.2018.1530197

Grégoire, L., and Greening, S. G. (2020). Fear of the known: semantic generalisation of fear conditioning across languages in bilinguals. *Cognit. Emot.* 34, 352–358. doi: 10.1080/02699931.2019.1604319

Hagemann, D., Naumann, E., Maier, S., Becker, G., Lürken, A., and Bartussek, D. (1999). The assessment of affective reactivity using films: validity, reliability and sex differences. *Personal. Individ. Differ.* 26, 627–639. doi: 10.1016/S0191-8869(98)00159-7

Haraway, D. (1988). Situated knowledges: the science question in feminism and the privilege of partial perspectives. *Fem. Stud.* 14, 575–599. doi: 10.2307/3178066

Harding, S. (1991). *Whose Science? Whose Knowledge? Thinking From Women's Lives.* Ithaca, New York (United States): Cornell University Press.

Haslanger, S. (2000). "Feminism in metaphysics: managing the natural" in *The Cambridge Companion to Feminism in Philosophy*. eds. M. Fricker and J. Hornsby (Cambridge (UK): Cambridge University Press), 107–126.

Hicks, M. (2017). *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing.* Cambridge, MA (USA): MIT Press.

Hodes, R. L., Cook, E. W., and Lang, P. J. (1985). Individual differences in autonomic response: conditioned association or conditioned fear? *Psychophysiology* 22, 545–560. doi: 10.1111/j.1469-8986.1985.tb01649.x

Igartua, J. J., and Muñiz, C. (2008). Identificación con los personajes and disfrute ante largometrajes de ficción. Una investigación empírica [Character identification and enjoyment of fiction feature films. An empirical investigation]. *Comunicacion Soc.* 21, 25–51.

Israelashvili, J., Oosterwijk, S., Sauter, D., and Fischer, A. (2019). Knowing me, knowing you: emotion differentiation in oneself is associated with recognition of others' emotions. *Cognit. Emot.* 33, 1461–1471. doi: 10.1080/02699931.2019.1577221

Izard, C. E. (2016). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspect. Psychol. Sci.* 2, 260–280. doi: 10.1111/j.1745-6916.2007.00044.x

Kentner, A. C., and Grace, S. L. (2017). Between mind and heart: sex-based cognitive bias in cardiovascular disease treatment. *Front. Neuroendocrinol.* 45, 18–24. doi: 10.1016/j.yfrne.2017.02.002

Koelstra, S., Mühl, C., Soleymani, M., Jong-Seok, L., Yazdani, A., Ebrahimi, T., et al. (2012). DEAP: a database for emotion analysis: using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Lang, P. J. (1980). "Behavioral treatment and bio-behavioral assessment: computer applications" in *Technology in Mental Health Care Delivery Systems*. eds. J. B. Sidowski, J. H. Johnson and T. A. Williams (New York City, New York (USA): Ablex Publishing Corporation), 119–137.

Lang, P. J. (1985). "The cognitive psychophysiology of emotion: Fear and anxiety," in *Anxiety and the Anxiety Disorders*. eds. A. H. Tuma and J. D. Maser (Lawrence Erlbaum Associates, Inc.), 131–170.

Lang, P. J., Bradley, M., and Cuthbert, B. N. (2008). International affective picture system (IAPS): instruction manual and affective ratings. (Technical Report A-8). The Center for Research in Psychophysiology, University of Florida. Available at: https://csea.phhp.ufl.edu/Media.html

Leavy, S. (2018). Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning in "*2018 IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 2018*," 14–16

Leen-Feldner, E. W., Blumenthal, H., Babson, K., Bunaciu, L., and Feldner, M. T. (2008). Parenting-related childhood learning history and panic vulnerability: a test using a laboratory-based biological challenge procedure. *Behav. Res. Ther.* 46, 1009–1016. doi: 10.1016/j.brat.2008.06.002

Lockwood, P. L., Ang, Y.-S., Husain, M., and Crockett, M. J. (2017). Individual differences in empathy are associated with apathy-motivation. *Sci. Rep.* 7, 17293–17210. doi: 10.1038/s41598-017-17415-w

Mauss, I. B., and Robinson, M. D. (2009). Measures of emotion: a review. *Cognit. Emot.* 23, 209–237. doi: 10.1080/02699930802204677

Miranda, J. A., Canabal, M. F., Gutiérrez-Martín, L., Lanza-Gutiérrez, J. M., Portela-García, M., and López-Ongil, C. (2021). Fear recognition for women using a reduced set of physiological signals. *Sensors* 21, 1587. doi: 10.3390/s21051587

Miranda, J. A., Rituerto-González, E., Luis-MinguezaCanabal, C., Canabal, M. F., Ramírez Bárcenas, A., Lanza-Gutiérrez, J. M., et al. (2022). Bindi: affective internet of things to combat gender-based violence. *IEEE Internet Things J.* doi: 10.1109/JIOT.2022.3177256

Miranda-Correa, J. A., Abadi, M. K., Sebe, N., and Patras, I. (2018). AMIGOS: A dataset for affect, personality and mood research on individuals and groups. *IEEE Trans. Affect. Comput.* 12, 479–493. doi: 10.1109/TAFFC.2018.2884461

Moltó, J., Segarra, P., López, R., Esteller, À., Fonfría, A., Pastor, M., et al. (2013). Adaptación española del "international affective picture system" (IAPS): Tercera parte [Spanish adaptation of the "international affective picture system" (IAPS): part 3]. *Anal. Psicol.* 29, 965–984. doi: 10.6018/analesps.29.3.153591

Moss-Racussin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., and Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proc. Nat. Acad. Sci. U.S.A.* 109, 16474–16479. doi: 10.1073/pnas.1211286109

Noble, S. (2018). *Algorithms of Oppression: How Search Engines.* New York City, New York (USA): NYU Press.

Nurock, V. (2020). ¿Puede prestar cuidados la Inteligencia Artificial? [Can artificial intelligence provide care?]. *Cuadernos Relac. Lab.* 38, 217–229. doi: 10.5209/crla.70880

O'Neil, C. (2016). *Weapons of Math Destruction: How big data Increases Inequality and Threatens Democracy.* New York City, New York (USA): Crown Random House.

Olatunji, B. O., Wolitzky-Taylor, K. B., Babson, K. A., and Feldner, M. T. (2009). Anxiety sensitivity and CO_2 challenge anxiety during recovery: differential correspondence of arousal and perceived control. *J. Anxiety Disord.* 23, 420–428. doi: 10.1016/j.janxdis.2008.08.006

ONU Mujeres (2021). Incorporación de la perspectiva de género. Available at: https://www.unwomen.org/es/how-we-work/un-system-coordination/gender-mainstreaming#:~:text=La%20Cuarta%20Conferencia%20Mundial%20sobre,compromisos%20en%20igualdad%20de%20g%C3%A9nero

Ortega-Toro, E., Jiménez-Egido, J. M., Palao-Andrés, J. M., and Sainz-de-Baranda, A. P. (2008). Diseño and validación de un cuestionario para valorar las preferencias and satisfacciones en jóvenes jugadores de baloncesto [Design and validation of a questionnaire to assess preferences and satisfaction in young basketball players]. *Cuadern. Psicol. Deporte* 8, 39–58.

Pedrosa, I., Suárez-Álvarez, J., and García-Cueto, E. (2013). Evidencias sobre la validez de contenido: avances teóricos and métodos para su estimación [Evidence on content validity: theoretical advances and methods for its estimation]. *Acción Psicol.* 10, 3–18. doi: 10.5944/ap.10.2.11820

Penfield, R. D., and Giacobbi, P. R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Meas. Phys. Educ. Exerc. Sci.* 8, 213–225. doi: 10.1207/s15327841mpee0804_3

Picard, R. W. (1995). Affective computing. M.I.T media laboratory perceptual computing section. *Tech. Rep.* 321, 1–26.

Picard, R. W., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: analysis of affective physiological state, IEEE T. *Pattern Analy. Machine Intellig.* 23, 1175–1191. doi: 10.1109/34.954607

Plutchik, R. (2001). The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am. Sci.* 89, 344–350. doi: 10.1511/2001.4.344

Revi Sterling, S. (2013). "Designing for trauma: the roles of ICTD in combating violence against women (VAW)." in *ICTD '13: Proceedings of the Sixth International Conference on Information and Communications Technologies and Development.* eds. G. Marsden, J. May, M. Chetty, and U. Rivett; 2, 159–162, Association for Computing Machinery.

Rituerto-González, E., Mínguez-Sánchez, A., Gallardo-Antolín, A., and Peláez-Moreno, C. (2019). Data augmentation for speaker identification under stress conditions to combat gender-based violence. *Appl. Sci.* 9, 2298. doi: 10.3390/app9112298

Rituerto-González, E., Miranda, J. A., Canabal, M. F., Lanza-Gutiérrez, J. M., Peláez-Moreno, C., and López-Ongil, C. (2020). "A hybrid data fusion architecture for BINDI: A wearable solution to combat gender-based violence" in *Multimedia Communications, Services and Security*. eds. A. Dziech, W. Mees and A. Czyżewski (Cham: Springer), 223–237.

Robinson, D. L. (2008). Brain function, emotional experience and personality. *Neth. J. Psychol.* 64, 152–168. doi: 10.1007/BF03076418

Sainz-de-Baranda, A. C., Blanco-Ruiz, M., Miranda, J. A., Gutiérrez-Martín, L., Canabal, M. F., San-Segundo, R., et al. (2021a). Perspectiva de género and social en las STEM: La construcción de sistemas inteligentes para detección de emociones [Gender and social perspective in STEM: Building intelligent systems for emotion detection]. *Soc. Technosci.* 11, 83–115. doi: 10.24197/st.Extra_1.2021.83-115

Sainz-de-Baranda, A. C., Blanco-Ruiz, M., and San-Segundo, R. (2021b). El rol del audiovisual en la activación de la memoria autobiográfica en víctimas de violencia de género [The role of audiovisuals in the activation of autobiographical memory in victims of gender violence]. *Cuestion. Género* 16, 810–835. doi: 10.18002/cg.v0i16.6918

Sainz-de-Baranda, A. C., De-Lamo Velado, I., and Nieto Rojas, P. (2022). Use of technological devices and (re)victimization in gender-based crimes in Spain: a qualitative study on professionals' perceptions. *Soc. Technosci.* 12, 56–72. doi: 10.24197/st.1.2022.56-72

Saleh, W., Leva, M. C., Ababio-Donkor, A., and Thimnu, A. (2022). "Gender and equality in transport." in *Proceedings of the 2021 Travel Demand Management Symposium*.

San-Segundo, M. R., Sainz-de-Baranda, A. C., Blanco-Ruiz, M., Larrabeiti-López, D., Urueña, P. M., Robledo, G. J. C., et al. (2021). Modelo de utilidad N° de solicitud. U202130953. Sistema y método para determinar un estado emocional de un usuario [System and method for determining an emotional state of a user]. Available at: https://bopiweb.com/2021-05-17.779#bopi_3753503

Schaefer, A., Nils, F., Sanchez, X., and Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers. *Cognit. Emot.* 24, 1153–1172. doi: 10.1080/02699930903274322

Schiebinger, L. (2021). Gendered innovations: integrating sex, gender, and intersectional analysis into science, health and medicine, engineering, and environment. *Tapuya Latin Am. Sci. Technol. Soc.* 4, 1867420. doi: 10.1080/25729861.2020.1867420

Soares, A. P., Pinheiro, A. P., Costa, A., Frade, S., Comesaña, M., and Pureza, R. (2013). Affective auditory stimuli: adaptation of the international affective digitized sounds (IADS-2) for European Portuguese. *Behav. Res. Methods* 45, 1168–1181. doi: 10.3758/s13428-012-0310-1

Soleymani, M., Pantic, M., and Pun, T. (2012). Multimodal emotion recognition in response to videos. *IEEE Trans. Affect. Comput.* 3, 211–223. doi: 10.1109/T-AFFC.2011.37

Sonderegger, A., Heyden, K., Chavaillaz, A., and Sauer, J. (2016). "AniSAM and AniAvatar: animated visualizations of affective states." in *The 2016 CHI Conference.* eds. J. Kaye, and A. Druin, 4828–4837.

Soto-Sanfiel, M. T., Aymerich-Franch, L., and Ribes-Guàrdia, X. (2010). Impacto de la interactividad en la identificación con los personajes de ficciones. *Psicothema* 22, 822–827. PMID: 21044519

Sumartojo, S., Pink, S., Lupton, D., and LaBond, C. H. (2016). The affective intensities of datafied space. *Emot. Space Soc.* 21, 33–40. doi: 10.1016/j.emospa.2016.10.004

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

Thaler, A. (2022). "Saving lives with gender studies? Putting technofeminism into practice" in *Proceedings of the 4th International Conference on Gender Research.* (eds.) E. T. Pereira, C. Costa, and Z. Breda.

Thomas, J. R., and Nelson, J. K. (2007). *Métodos de investigación en actividad física [Research methods in physical activity]*. Barcelona: Paidotribo

Trommsdorff, G., Friedlmeier, W., and Mayer, B. (2016). Sympathy, distress, and prosocial behavior of preschool children in four cultures. *Int. J. Behav. Dev.* 31, 284–293. doi: 10.1177/0165025407076441

Verma, G. K., and Tiwary, U. S. (2015). Affect representation and recognition in 3D continuous valence–arousal–dominance space. *Multimed. Tools Appl.* 76, 2159–2183. doi: 10.1007/s11042-015-3119-y

Wajcman, J. (2006). El tecnofeminismo [Technofeminism]. Universitat de Valéncia.

Wiersma, L. D. (2001). Conceptualization and development of the sources of enjoyment in youth sport questionnaire. *Meas. Phys. Educ. Exerc. Sci.* 5, 153–177. doi: 10.1207/S15327841MPEE0503_3

Zafra, R. (2011). Un cuarto propio conectado. Feminismo and creación desde la esfera público-privada online [A connected room of one's own. Feminism and creation from the online public-private sphere]. *Asparkía Investig. Femin.* 22, 115–129.

Zaki, J., and Ochsner, K. N. (2015). "Empathy," in *Handbook of Emotion*. eds. B. L. Feldman, M. Lewis and J. M. Haviland-Jones (New York City, New York (USA): Guilford), 1–34.

frontiers | Frontiers in Psychology

TYPE Conceptual Analysis
PUBLISHED 18 November 2022
DOI 10.3389/fpsyg.2022.960327

Check for updates

OPEN ACCESS

EDITED BY
Milagros Sainz,
Open University of Catalonia, Spain

REVIEWED BY
Itziar Fernández,
National University of Distance Education, Spain
María Cecilia Fernández DArraz,
Temuco Catholic University, Chile

*CORRESPONDENCE
Lydia Bradford
bradf134@msu.edu

SPECIALTY SECTION
This article was submitted to Gender, Sex and Sexualities, a section of the journal Frontiers in Psychology

RECEIVED 02 June 2022
ACCEPTED 27 September 2022
PUBLISHED 18 November 2022

CITATION
Schneider B, Chen I-C, Bradford L and Bartz K (2022) Intervention initiatives to raise young people's interest and participation in STEM.
Front. Psychol. 13:960327.
doi: 10.3389/fpsyg.2022.960327

Intervention initiatives to raise young people's interest and participation in STEM

Barbara Schneider, I-Chien Chen, Lydia Bradford* and Kayla Bartz

College of Education, Michigan State University, East Lansing, MI, United States

For nearly a decade, two science interventions anchored in project-based learning (PBL) principles have been shown to increase student science learning in 3rd grade and high school physical science classes. Both interventions employed a randomized control trial of several thousand students (N=3,271 in 3rd grade and N=4,238 in 10th, 11th, and 12th grades). Incorporating a rich background of research studies and reports, the two interventions are based on the ideas of PBL as well as the National Academies of Science's publications, including how children learn; how science learning and instruction can be transformed; and the performance expectations for science learning articulated in the Next Generation of Science Standards. Results show significant positive increases in student academic, social, and emotional learning in both elementary and secondary school. These findings can be traced, in part, to carefully crafted experiential participatory activities and high-quality instructional materials which act as strong facilitators for knowledge acquisition and use. Reviewing the innovations undertaken by these two interventions, this article describes the importance of studying social and emotional factors *'in situ'*, using the Experience Sampling Method (ESM), that can motivate and engage students in science learning in both elementary and secondary school. Using these *'in situ'* data collection (N=596 students in 3rd and N=1412 students in 10th, 11th, and 12th grades) along with case studies and repeated measures analysis gave deep insights into emotional and social development for young children and adolescents. These methods should continue to be considered when trying to understand key factors of improving engagement in science.

KEYWORDS

engagement, social and emotional learning, science learning, interventions, project-based learning

Introduction

National and international assessments indicate that US students' academic performance in science is barely reaching average scores, especially in junior and senior high school (National Assessment of Educational Progress, 2021; Organisation for Economic Co-operation and Development, 2020). More disconcerting is that among certain segments of the US student population performance scores continue to lag behind most students in the general population. Additionally, the stagnant or marginal declining scores of 4th graders on NAEP in 2019 and no changes in the scores of 12th graders affirms the view of researchers, business community, and public stakeholders that US students are unprepared to meet the technological changes of today and likely to have difficulty finding stable employment as adults (see Hammerstein et al., 2021).

These less than promising science achievement test results were evident before the SARS-CoV-2 pandemic. The latest projections, especially among those with the most limited economic and social resources, is that these students are likely to experience major academic, social, and emotional problems at school this coming year and perhaps throughout their careers and beyond (Dorn et al., 2020). The pandemic has raised multiple questions about the long-term effects on student lives and their resilience, having experienced an unprecedented global health crisis. One possible solution for ameliorating these long-term effects is implementing new promising interventions with innovative instructional strategies and materials which show results of increasing science achievement as well as the social and emotional needs of children and adolescents.

For the past several years, two science curriculum interventions have been implemented and evaluated in elementary and secondary schools (Schneider et al., 2022; Krajcik et al., in press). These two grade level interventions share a theoretical design-based rationale, based on project-based learning (PBL) principles, and provide solutions to several serious questions that have been raised about the quality of science instruction in the US. The elementary school intervention, Multiple Literacies in Project-based Learning (ML-PBL), is an efficacy study of 3rd graders in Michigan where students were given four science units in which learning goals were developed consistent with the Next Generation of Science Standards and the instructional experiences were based on the components of three-dimensional learning (disciplinary core ideas, science and engineering practices, and cross-cutting concepts; National Research Council, 2012; Krajcik et al., 2021). The high school study, Crafting Engaging Science Environments (CESE), developed three units in chemistry and three in physics, and was similarly created using the PBL principles, the NGSS performance expectations, and the National Research Council's definition of three-dimensional learning (Schneider et al., 2022). These units, at the elementary and secondary level, were all designed with experiences to promote students asking questions, collaborating with one another, constructing evidence and artifacts, and engaging in scientific and engineering practices.

One key, new addition found in the ML-PBL and CESE interventions was the explicit importance placed on social and emotional learning and its relationship to science achievement (see chapter by Krajcik and Schneider, 2021 for ML-PBL; see Schneider et al., 2020 for CESE). PBL has implicitly emphasized social and emotional learning with its activities, materials, and assessments that have been deliberately designed to create equitable environments (Miller and Krajcik, 2019). In PBL classrooms all students are encouraged and supported to participate in asking questions, collaborate and work in teams, and share personal science experiences both in- and out-of-the classroom. However, in these two interventions, these ideas were further articulated theoretically and applied with specific methods and items developed to measure the impact of social and emotional factors on science achievement at the elementary and secondary levels. At the elementary level, students were asked questions about their interest, skills, and challenge in specific science activities (Bartz et al., 2022) and these same measures were asked to the secondary students with age-appropriate language (Schneider et al., 2016; Bradford and Bartz, 2022). Additionally, during teacher professional learning sessions, special activities were designed to guide teachers in fostering greater participation and inclusivity among all students (Krajcik and Schneider, 2021; Schneider et al., 2022).

Several considerations in the design of the interventions were identified for understanding social and emotional learning for both elementary and secondary students. First, and most importantly, was the selection of social and emotional constructs that were appropriate for science learning in classrooms (Baines et al., 2017; Lee et al., 2019). Care was taken not to include the entire corpus of psychological measures of emotionality but rather concepts that could be observed (i.e., self-reflection, ownership, and collaboration) and assessed during science lessons. Second, because the focus was on promoting engagement in science learning, fundamental concepts identified in earlier studies of engagement were used to measure interest, skills, and challenge when involved in learning activities (Salmela-Aro et al., 2016; Schneider et al., 2016; Moeller et al., 2017).

For purposes of measurement at the elementary level, social and emotional learning states (i.e., patterns of feelings during activities within specific time periods) were assessed when students were in their science classes. This was a one-time measure, validated through a variety of statistical procedures (see, Krajcik et al., 2021). At the high school level, these constructs were measured '*in situ*' when students who were participating in PBL experiences multiple times during the semester, were randomly notified and asked to answer a survey on their emotionality with the Experience Sampling Method (ESM). The ESM is a type of time diary which uses repeated measures randomly obtained through an intermediate notification system (such as on a phone). The results from these two important additions to the PBL design showed significant positive impacts on science learning, motivation, and engagement. This chapter describes why social and emotional learning is an essential component for academic learning, how we incorporated them in these two different efficacy studies, and how we plan to evaluate their impact on science learning.

Why we need to care about social and emotional learning

More recently, there has been increased attention within the psychological community to investigate the relationship between the impact of social and emotional learning on student performance in classrooms. Previously, these issues were rarely isolated to the learning context or used to direct teachers' practices in their classrooms for the purpose of supporting all students' academic performance and well-being (Durlak et al., 2015; Jagers et al., 2018; Lee et al., 2019). This increased interest in contextualized social and emotional learning support has now been expanded in multiple frameworks to include sensitivity to differences in students' cultures, equity practices to encourage student participation in classroom experiences, and opportunities for enhanced collaborative and team activities [more specifically as discussed in Lee et al., 2019 and National Academies of Sciences, 2021].

The intentionality of inclusionary social and emotional learning opportunities in classrooms complements PBL principles (Peele-Eady and Moje, 2020; Rosado-May et al., 2020) and the execution in the design of ML-PBL and CESE interventions. One of the most critical aspects of PBL is beginning with a "driving question," a real-world problem, where students are encouraged and supported to ask meaningful questions that personalize the lesson to their own lives. Based on the driving question, subsequent experiences are enacted whereby students work together finding solutions to these problems over the course of a unit. The significance of the driving question is critical for motivating interest from the perspective of the students' lives, shaped by their familial and community economic, social, and cultural resources, and forging them on a path of personalized scientific inquiry and discovery. One cannot overlook the fundamental value of beginning science lessons from the standpoint of appealing to the personal interest of the students for "why" pursuing a recognizable puzzling phenomenon in their natural world may have importance to them (Renninger and Hidi, 2020). Results show that personalized meaningful interest in a topic motivates sustained interest in other science learning experiences providing that they are reasonable for the students' skill sets and are challenging solvable problems. By incorporating these ideas, students are more likely to persist and learn phenomena they may have previously considered unsolvable (Csikszentmihalyi and Schneider, 2000).

The ML-PBL and CESE interventions included carefully crafted lessons which are planned with a series of intra- and inter-connected experiences which coherently increase in scientific knowledge and practices (Fortus and Krajcik, 2012). Lessons are constructed so students complete them with their classmates or individually extend their competencies in planning investigations, building models, and writing scientific explanations, all of which offer support for learning how to formulate evidential claims to problems. Activities typically focus on "hands on" experiences, most often in groups, bringing together students of initially varying ability to have the opportunity to acquire actual scientific skills. These learning experiences are quite different from traditional science instruction which tends to rely on science content that students have to memorize, frequently measured individually with summative tests, and which frustrates many students contributing to the loss of interest in science (National Research Council, 2012; National Academies of Sciences, 2021). Rather, these sequential learning experiences are designed to challenge students to work on problems to which they do not know the answer and to encourage them to continue trying to solve them. These activities, which push students to seek the answer to challenging questions, while doing something important to them, have been shown to be related to feelings of determination (Bradford and Bartz, 2022).

Concentrating on several of the most important social and emotional learning measures, these two interventions also underscored the importance of obtaining such information on these constructs when students are in their science classes. This led in both intervention studies to several assumptions regarding social and emotional measurement: (1) SEL is not a distinctive single psychological state, one can be engaged and feel successful and in control while also feeling a sense of stress; (2) SEL is time variant, in that a confluence of SEL states vary in intensity across the course of one's daily life experiences; and (3) SEL is highly susceptible to contextual environmental conditions such as the instructional activities in the classroom.

Recognizing developmental differences in literacy, social and emotional awareness of self and others, and technological skills (Lerner and Steinberg, 2009), the selection of SEL measures and the methodology used in the elementary and secondary intervention studies varied in form, rapidity, and replication. However, what they shared is an overlap of SEL states that examined interest, sense of self-appraisal of one's involvement in specific activities, value of one's accomplishments, and collaboration with one another. The elementary design was to measure SEL during their science classes. The secondary school study examined moment to moment '*in situ*' experiences of when students were both within and outside their science classes which allowed for the measuring of variations in engagement, its construct validity, and its relationship to academic performance.

Study 1

Beginning as a design-based study for 3rd grade, the ML-PBL intervention underwent several rounds of revisions and testing over the course of 4 years, including teacher experiments, classroom pilots, a field-test, and most recently an efficacy study to determine whether the ML-PBL intervention enhanced students' science academic, social, and emotional learning. A randomized control trial was conducted in 46 Michigan schools (23 treatment and 23 control) which included four regions in the state. The final analytic sampled included a total of 2,371 students. The treatment condition included curriculum materials and professional learning experiences for teachers. To assess if there was a significant difference in academic science learning, a three-level hierarchical linear model (HLM) was conducted. This method was used to account for nesting of students within classrooms within schools. Results showed that the treatment

10.3389/fpsyg.2022.960327

students outperformed the control students by a .277 standard deviation on an objective summative test which is a substantial treatment effect (Krajcik et al., in press). This could be interpreted as a ten-point increase on a hundred-point scale or based upon a chosen percentile ranking in which the treatment could move the student from below proficient to proficient (Kraft, 2020).

The above work also investigated specific research questions related to social and emotional learning, specifically, whether the treatment support more positive responses on measures in self-reflection, collaboration, and responsibility for their own and others' work. It is important to underscore that few studies measure elementary school students social and emotional learning in their science classes (National Research Council, 1999). Given these constraints, the team consulted relevant limited science studies of young children and more broadly: psychological research studies on SEL; developmentally appropriate questions for 3rd graders; and items from other national assessments (e.g., the Early Childhood Longitudinal Study [ECLS-K] 2016; Durlak et al., 2015; Baines et al., 2017; Jagers et al., 2018). Recognizing differences in literacy skills among students, a drawn thumbs-up (agree), thumbs-down (disagree), and closed fist (neutral) were used to measure agreement. Students circled their feelings on a paper/pencil form administered in spring semester. Prior to the efficacy study, the SEL instrument was designed, field-tested, and revised. Confirmatory factor analyses (CFA) were performed and supported three key latent constructs: self-reflection, ownership, and collaboration (see Krajcik et al., 2021). Additionally, the reliabilities of these constructs were estimated to be 0.78 for self-reflection, 0.81 for ownership, and 0.74 for collaboration. Results from the efficacy study of the ML-PBL intervention showed that the treatment students were estimated to have 0.544 higher factor scores in reflection, 0.434 higher factor scores in ownership, and 0.416 higher factor scores in collaboration than the control students (Krajcik et al., 2021). These results indicate that it is possible to obtain validated measures of young children's SEL responses for selected constructs. And in this instance, constructs that are specifically designed to be contextually relevant for particular SELs that the intervention was expected to impact.

As mentioned above, few studies have been able to examine the impact of engagement on elementary science learning. We chose to further examine the relationship between engagement and achievement as research has shown positive relationships between students' determination to be engaged in the classroom and science achievement (Grabau and Ma, 2017). How students are feeling at the time of the lesson or activity can play a major role in how well they learn or understand key concepts. To explore student responses to project-based and three-dimensional learning, we developed optimal learning surveys that allowed us to measure student engagement in a repeated measures design. These surveys obtained student responses '*in situ*' within the science classroom, capturing students' perspectives on specific lessons as they happen (Csikszentmihalyi and Larson, 2014). Results of this new development study are described below.

During the beginning and first year of the Covid-19 pandemic, the ML-PBL team were able to observe students in their science classes *via* video, in-person, or a mixture of the two to collect data on the students' engagement and teacher implementation. One of major observations from the videos was the variation of students' engagement in their science lessons. Having identified in earlier studies of secondary students a set of constructs (i.e., interest, skill, and challenge) that showed increases in engagement and impacted science learning (see Schneider et al., 2016, 2020), the research team decided to pilot whether these same engagement constructs could be found in elementary science classes and whether they might also positively influence students' science learning.

Research questions

The research questions for this new development study include:

1. Can 3rd graders reliably produce measures of interest, skill, and challenge '*in situ*'?
2. When studied with repeated measures, do interest, skill, and challenge load onto a single construct of engagement?

Method

Using the same constructs of interest, skill, and challenge as fundamental dimensions of engagement, during the pandemic, the team developed a new methodology and series of items for 3rd graders that relied on data collected situated in specific lessons within each unit. Keeping with the idea of measuring social and emotional learning '*in situ*,' specific items were contextualized to be consistent with the lesson learning goals and how teachers may have been adapting them in the four units (see Bartz et al., 2022).

Instruments/measures

For each unit during three different time periods, students were asked questions pertaining to specific measures of interest, skill, and challenge (see page 7 for fuller description). The three different time points were chosen based on the goals of each lesson, allowing us to collect more data from lessons that focused specifically on driving questions, investigation, building a model, or creating a final artifact. These items are situated directly in the context of each lesson. Six focal lessons, which contained the following features: driving question, modeling, investigation, and development of a final artifact, were sampled. For example, in the beginning of the toy unit after observing a toy rocket and how it moves, students were asked for interest, "I like asking questions about how the air rocket moves;" for skills, "I can ask questions about how toy rockets move the way they do;" and for challenge, "I had to think a lot to ask new questions about how rockets move." With respect to collaboration,

TABLE 1 Sample descriptives.

	N	Mean	St. Dev	Min	Max
Interest	3,369	3.29	0.87	0	4
Skill	3,330	3.25	0.84	0	4
Challenge	3,367	2.79	1.08	0	4
Q4	3,362	3.18	1.01	0	4

TABLE 2 Reliability using Cronbach's Alpha.

	Item-test	Item-rest	Avg. interitem cov	Alpha
Interest	0.67	0.39	0.17	0.4
Skill	0.62	0.34	0.2	0.44
Challenge	0.61	0.21	0.25	0.56
Q4	0.69	0.35	0.17	0.42
Test scale			0.2	0.53

the students were asked, "When I worked with my classmates, we came up with different questions about the way the toy rocket moved;" and for ownership, "The questions I asked about the air rocket's motion were important to me and my classmates."

The data collection procedures used for measuring this engagement measure followed the original collection of the SEL survey, but with greater frequency. Teacher administered the four-question OLM survey to third grade students immediately following the lesson. The first three questions were based on engagement: interest, skill, and challenge. The fourth varied by form (A, B, or C) and rotated between collaboration, persistence, agency, time and outcome by lesson. A 4-point Likert scale was used (strongly disagree, disagree, agree, strongly agree) with students circling icons of thumbs up and thumbs down. In the pilot of the SEL measures at the elementary school level, at three different times during each of the four units, the teachers hand out paper copies of the engagement questions to the students in their class. The teachers then read aloud each of the questions, one at a time. After each question is read, students circle the corresponding thumb icon on their paper. In the cases where students circled more than one response, in the median score of responses was recorded.

Sample

The sample for this analysis came from 25 3rd grade classrooms in Michigan and included 596 students with a total of 3,369 responses for an average of 6 repeated measures per student.

Analysis

Their responses to the engagement questions across the four ML-PBL units were analyzed. For the reliability of this survey, a Cronbach's alpha was used to estimate the reliability.

$$alpha=\left(\frac{K}{(K+1)}\right)\left(1-\frac{\sum Vi}{Vt}\right).$$

For understanding whether the interest, skill, and challenge loaded onto a construct of engagement, a confirmatory factor analysis was conducted. Factor loadings for each item onto this construct were estimated.

Results

The descriptive statistics from the survey, including the items of interest, skill, challenge, and an additional question, are reported in Table 1.

A confirmatory factor analysis confirmed a unidimensional model with the following factor loadings for: interest (0.77); skill (0.41); and challenge (0.26). The overall reliability of the engagement measure is a Cronbach's Alpha of 0.53. The overall reliability and item level reliabilities are reported in Table 2.

Additional analyses are being undertaken to study variation in engagement by lesson activities and individual level variables.

Study 2

The secondary school intervention, "Crafting Engaging Science Environments," (CESE) is a high school chemistry and physics PBL intervention similar to but independent of the elementary intervention. Both interventions meet the NGSS performance expectations and incorporate NRC three-dimensional learning and principles of PBL. CESE was administered to a diverse group of over 4,238 students in chemistry and physics classes in 70 high schools. The design like the elementary study was an efficacy study that involved a randomized control trial in California and Michigan. This intervention also included curriculum materials and professional learning for the teachers. Results were estimated using a two-level HLM with the outcome being the student level performance on the physical science items from the Michigan State Science Assessment and the main predictor of interest being treatment at the school level. For this estimation, a pretest and student demographics were included as covariates. Results show that treatment students, on average, performed 0.20 standard deviations higher than control students on an independently developed summative science assessment (Schneider et al., 2022). These results, like the ML-PBL, are quite large especially considering the advanced subject matter of the units and that they only extended over a 12-to-16-week period. Mediation analyses show an indirect path between teacher- and student-reported participation in modeling practices and science achievement. Exploratory analyses, using a two-level mixed logit model also indicate positive treatment effects for enhancing college ambitions. Overall, results show that improving secondary school science learning is achievable with a coherent system comprising teacher and student learning experiences, professional learning, and formative unit assessments that support students in "doing" science.

A major part of the study was investigating why secondary students, as shown in national and international studies fail to be engaged in their science classes which likely affects their interest in science learning, achievement, and science career ambitions (National Assessment of Educational Progress, 2021; Organisation for Economic Co-operation and Development, 2020). This question of how to enhance engagement in science was a major concern of the secondary school science study. Several major hypotheses about studying engagement were assumed at the onset of the study as discussed above that students' social and emotional experiences at school are fluid throughout their daily lives (Csikszentmihalyi and Schneider, 2000). First, as discussed above students' social and emotional experiences at school are fluid throughout their daily lives. It is not expected that students would be fully engaged in all their classes full-time any more than it is expected that adults would be consistently fully engaged in all activities at work or at home. Moreover, because of what is known about adolescent development, trying to create activities that keep teenagers fully engaged requires quite a high bar of motivation (Immordino-Yang, 2015). Irrespective of the barriers and challenges, the problem to be addressed in this study was creating environments that were engaging. The nature of science requires inquiry-based discovery (National Research Council, 2012; National Academis of Sciences, 2018); therefore, students may be more receptive to doing science than memorizing facts or plugging in equations.

The PBL framework, which stresses solving personally meaningful questions and encouraging instructional activities that require collaboration and are intellectually challenging, was ideally suited to test the constructs of engagement and their impact on academic science achievement. The work is situated in the work of Fredricks and McClolskey (2012) that identifies engagement as having cognitive, behavioral, and subjective components. Extending their definition, the new conception of engagement begins by identifying special behavioral activities that are temporal in quality, spark personalized interest, require competence of a set of knowledge and experiential science practices, and undertake challenging problems.

In contrast to those who have conceptualize engagement as a general trend, this model of engagement identifies engagement as domain specific in duration and in intensity, which fits more closely with current definitions of situational interest in science learning (see Lavonen et al., 2005; Krapp and Prenzel, 2011). This situational approach is different from other scholars who are interested in identifying universal traits (Deaux and La France, 1998; Cuddy et al., 2008). These engagement experiences are defined as optimal learning moments, which also builds upon the idea of "flow" defined by Csikszentmihalyi and Csikszentmihalyi (1988) as situation specific instances when an individual is so deeply involved in a specific task-related activity that time flies by (Csikszentmihalyi, 1990; Hektner et al., 2007).

The PBL curriculum, as discussed above, begins with a driving question when students are in specific situations and faced with a problem or phenomenon that is relevant and meaningful to their lives, such as: "how can I build a safer car?" To build that car, students need to have the necessary knowledge and skills to create a solution. Irrespective of the students' skill level, finding a reasonable solution should be a challenge, one that sparks determination. When students are fully engaged in a learning task, this is defined as an optimal learning moment (OLM). These moments do not just happen, but need to be artfully constructed and coherent, which is yet another fundamental aspect of PBL which inspires the acquisition of new knowledge, the use of imagination, and stretching problem-solving abilities.

Optimal learning moments can be verified and understood by other related subjective experiences occurring at nearly the same time. For example, it is expected that when involved in these activities' students feel successful, confident, active, happy, and enjoyment with the activity (Shernoff et al., 2003; Shumow and Schmidt, 2014). Learning accelerants are those experiences of feeling anxious or stressed, which activate learning (Csikszentmihalyi and Schneider, 2000). Finally, the contrast to positive subjective experiences, termed learning detractors, is when students involved in an activity feel confused or bored and are therefore less likely to be actively engaged or experience an OLM (Csikszentmihalyi, 1990; Schneider et al., 2016).

During the field test of the CESE intervention, an '*in situ*' study of social and emotional relationships to science achievement was conducted with the ESM. The data included 8,273 responses from 244 students in 15 classes taught by 14 teachers in Michigan. Only half of the variance in determination and giving up were at the student level, meaning that both feelings are not altogether stable student traits and most importantly environment and context matters (Schneider et al., 2020). Students were more likely to report giving up when tasks became more challenging, but at the same time, when classroom activities were reported as more challenging than average, students were more likely to persevere, suggesting that determination is partially situationally dependent and shaped by what is occurring in the types of activities presently involved in either with others or oneself.

While these ESM results were promising, there were several limitations. This was a pilot not a randomized trial where students in a treatment and control group could be compared. Rather it was the case that measures of engagement and feelings regarding challenge were measured using a single case design, where each classroom acts as its own experimental control (vacillating from treatment periods to times in the classroom when it was "business as usual"). These repeated periods were assessed to determine if the treatment influenced students' engagement. Although, the pilot study results showed that more engaged students had higher grades it could not be directly attributable the CESE intervention. However, the positive nature of the results prompted the team to use the ESM in the future efficacy study (2018–2019) in selected treatment and control classrooms (Schneider et al., 2020).

Preliminary results on the measures of engagement show that when considering levels of interest skill, and challenge, student engagement levels increase and are accompanied by other positive social and emotional affects, as well as decline in feeling of boredom and confusion. These findings show that concepts such as engagement, creativity, and problem-solving are situationally specific and share nearly equal variance when contrasted with person-level characteristics. In other words, even if a student is not interested in a topic or whose previous science achievement scores are below average, a carefully created situation can alter their negative predilections toward science, bringing considerable strength to the "nurture" side of learning especially when breaking from traditional types of assessment memorization and instead using imagination, problem-solving, and taking different points of view into consideration when engaged in scientific practices. However, these are preliminary results and an important question is the level of challenge and what impact it has on motivating higher engagement and learning for all students in specific contexts. (see, Schneider et al., 2020, for a deeper discussion of these ideas and how they were conceived and measured in the earlier field study).

Current study

Most recently, a deeper examination of "challenge experiences" in science class has been conducted (Bradford and Bartz, 2022; Chen et al., 2022). Until now, challenge has not been a major state in the psychological literature, and less attention has been placed on perceived "challenge" experiences in science classroom environments. Challenge experience can be highly motivating and encourage deeper engagement in a classroom task. However, there is less research regarding the importance of perceived challenge for high schoolers, how it varies '*in situ*', and how students react to challenging experiences. To fill the gap, this study has two different analyses.

The first starts with a *case study approach* to illustrate a particular pattern of perceived challenge by visualizing three student cases in 4 days of their school life. The visualization focuses on a precise moment in time and provides corresponding details on where students were, what they were doing, and who they were with. From there, the graphic visualization considers students' school life for 4 days and how this pattern of perceived challenge experiences is general or unique to individuals who vary in their background science knowledge. After visualizing three students' life in school, we use another graphic layout to visualize students' reactions to perceived challenges in their positive and negative states. The purpose of this visualization work is to lead the researchers to discover patterns of emotionality shared by several members of the student sample for 4 days.

The second analysis uses data from a sample of students from the field test and efficacy study to understand the use of ESM and student's variation in emotions across years. These analyses employ a series of repeated measures estimation of students situational perceived challenge, stress, anxiety, determination, giving up, and confusion to understand how the relationship between challenge and giving up and confusion is mitigated by stress and anxiety. We assume that challenge is important in driving learning; however, if challenge is correlated to giving up and confusion, this would lead to a negative relationship between challenge and learning. This leads to a question of whether anxiety and stress may be stronger mediating factors in the relationship between challenge and giving up and confusion.

Research questions

The research questions for study 2 were:

1. How does perceived challenge vary by individual students?
2. How does the relationship between perceived challenge and positive and negative emotions vary by individual students?
3. What is the relationship between students' perceived challenge coupled with stress and anxiety and determination, giving up, and confusion?

Sample

During the field test of the CESE (2013–2018), a total of 867 students were reported with the ESM. For the efficacy study (2018–19), a total 545 students were reported with the ESM for a total of 1,412 students combined. The phones were programmed to alert the students randomly 6–8 times per day (at least 3–4 times when they had science lessons) over an assigned period. An initial ESM prompt would occur in the beginning, mid- and late point of a study session automatically set up by researchers using the PACO app. Students were asked to respond to an identical questionnaire (nearly 30 items) within a 15 min window. Two reminders would occur 10 and 15 min after the initial prompt. On average, it takes about 90 s to complete items. Each day all participants received eight to 10 beeps on their smartphones which gave them 40 total response opportunities during a study period. We preprogrammed the beep schedule randomly and guarantee a minimum of 1–3 beeps occurring in science classes, resulting in 5 to 15 beeps per person in this study. In total, the data comprised 3,234 responses. The average valid beeps per student is 6 in science classes. We conducted two separate analyses, one which only analyzed the students in the efficacy study and a second analysis from both the field and efficacy studies.

The first analyses reported is from the efficacy study which contained a diverse population of students living in both Michigan and California with an overrepresentation of students for whom English is not their first language, as one of our sites was a mile from the Mexican border. Among the efficacy students' sample, 315 (58%) had valid student background information, including Race/Ethnicity, gender, challenge experiences and science pretest scores. This student background survey was collected at the beginning of the year *via* a Qualtrics Survey. Table 3 reports the descriptive statistics for the students. Of the 315 students who provided valid ESM responses, the racial composition of the

TABLE 3 Descriptive statistics of student sample.

	Freq.	%
Male	151	48.55
Female	160	51.45
Grade 10	81	31.89
Grade 11	146	57.48
Grade 12	27	10.63
White, (non-hispanic)	200	60.4
Hispanic	57	17.2
Black	28	8.5
Asian	18	5.4
Other	11	3.3
Multiracial	17	5.1
Total valid student info	331	
Demographic info Missing	214	
	Mean	SD
Percentile ranking of pre-test	62.69	22.03
Challenge	2.31	0.74

groups was 60% White, 8% Black, 17% Hispanic, 5% Asian and 5% multi-racial.

The second analysis used the entirety of the sample from both the field and efficacy tests (demographic information is unavailable for this combined sample; however, the sampling scheme for the field and efficacy tests targeted schools with significant numbers of low-income and minority students). The entire sample was used in the second analysis which uses aggregate statistical modeling to understand the validity of these relationships across many years (2013–2019).

Methods

Instruments and measures

To measure engagement, studies typically employ surveys which are rarely conducted '*in situ*' or when they are happening, which of course fails to capture how students are feeling from one moment to the next. Measuring how students feel across moments allows us to identify when they feel successful at what they are doing and its relationship to what they are learning. The ESM records what students are doing, what they are thinking about, and what they report feeling in the moment forming an archival repository of daily experiences. This focus on the situational and contextual aspects of what happens in-and -out-of-the classroom lessens the opportunity for recall bias and socially desirable answers and has been validated in previous studies (Hektner et al., 2007). The ESM SEL survey items and their response are reported in Table 4. There were approximately 30 items.

In both analyses, students were beeped several times a day (7 times) during a week both inside and out of school and classes, with several more signals in science classes. Each classroom was randomly chosen for a specific week(s) during the intervention

TABLE 4 CESE ESM instrument.

ESM questions in CESE
Q1 Where were you when you were signaled?
Q2 What science class were you in?
Q3 Which best describes what you were doing in science when signaled?
Q4 What were you doing when signaled?
Q5 What were you learning about in science when signaled?
Q6 Who were you with?
Q7 Were you doing the main activity because you…
Q8 Was what you were doing…
Q9 Were you interested in what you were doing?
Q10 Did you feel skilled at what you were doing?
Q11 Did you feel challenged by what you were doing?
Q12 Did you feel like giving up?
Q13 How much were you concentrating?
Q14 Do you enjoy what you are doing?
Q15 Did you feel like you were in control of what you were doing?
Q16 Were you succeeding?
Q17 Was this activity important for you?
Q18 How important is this activity in relation to your future goals/plans?
Q19 Were you living up to the expectations of others?
Q20 Were you living up to your expectations?
Q21 I was so absorbed in what I was doing that time flew.
Q22 How determined were you to accomplish the task?
Q23 Were you feeling…Happy
Q24 Were you feeling… Excited
Q25 Were you feeling… Anxious
Q26 Were you feeling… Competitive
Q27 Were you feeling… Lonely
Q28 Were you feeling… Stressed
Q29 Were you feeling… Proud
Q30 Were you feeling… Cooperative
Q31 Were you feeling… Bored
Q32 Were you feeling… Self-confident
Q33 Were you feeling… Confused
Q34 Were you feeling… Active

for data collection. Each data entry has a time stamp to indicate when the responses was collected. This approach is different from single survey as it records a set of repeated specific social and emotional measures interacting with specific activities, such as, doing a hands-on experiment in science class as compared to playing a video game. These responses are uploaded to a secured server which sends information to a cloud and are then quickly transformed into clean datasets and ready for analysis. Confidentiality is maintained by student anonymized identification numbers (It is important to note that all of our data collection and analyses underwent Institutional review board approval and received exempt status).

Figure 1 below shows a screen shot of one of the questions used in the secondary school intervention in both the field test and efficacy study. The actual software program was developed by

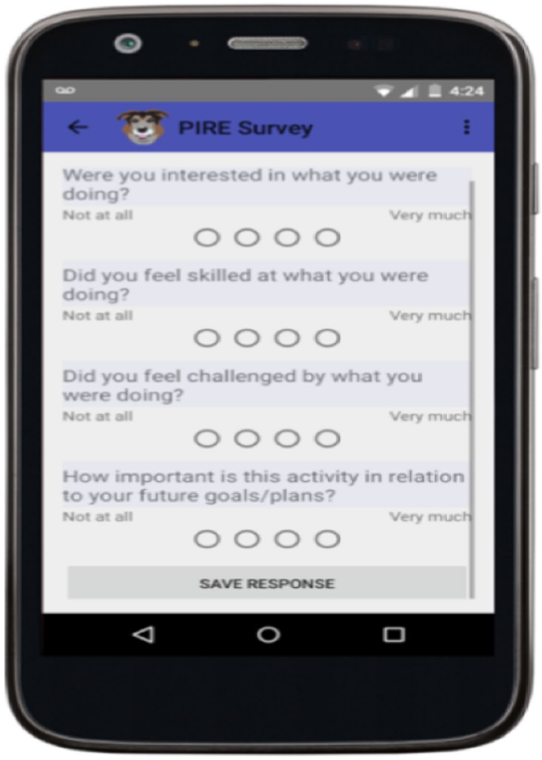

FIGURE 1
Screen shot of PACO app (Evans, 2016). Republished with permission.

Robert Evans, a google engineer, who named the program Paco,[1] after his dog which barks to let students know it is time to answer the questions. Although there are multiple questions, the students can move through them quickly. Since they are programmed for smartphones, beeping schedules can be easily programmed over the course of days of a week or multiple weeks during specific time periods. Figure 1 shows the questions asked regarding interest, skill, and challenge.

Analysis

A more in-depth examination of ESM is shown in the case study analysis. A second analysis which relies on a hierarchical linear model (HLM) was used for the aggregate study of the field and efficacy tests. Beginning with the in-depth case study, three students were selected among varying levels of school achievement to analyze their variability in emotionality with graphic visualization. First a description is given for how an individual student experienced challenges across activities, locations, and companionship in their 4 days. Second, the three students' emotional responses within each person's positive and negative states when challenged is shown in Figures 2–4. Finally in the second analysis, we explore the relationship of challenge on spurring continued determination or on confusion and giving up with or without changes in other states of emotionality through the HLM.

Three case studies

To understand the situational and individual differences for the students in the case study, their ESM responses were obtained throughout the day, including an oversample of beeps in their science classes (Chen et al., 2022). Table 5 shows the three students' background, the level of prior test scores, and the average perceived challenge across all contexts and in the science classroom only. To see how this visualization works, consider the rating of perceived challenge by those three students. The three students are: Dennis, a low-academic performing student based on his prior test scores; Megan, an academically average student; and Collins, an above average on his prior test scores. During the 4 days of the study, Dennis has an average challenge response of 3.14 across all contexts and the average challenge response of 3.25 in the science classroom. Megan has an average challenge response of 2.51 across all contexts and the average challenge response of 2.25 in her science classroom. Collins has an average challenge response of 2.29 across all contexts and the average challenge response of 2.30 in his science classroom. These three students are in the same science class at their school. Examining these individual case studies allows for the comparisons among the three students, their different social and emotional experiences throughout the day, and their relationship to challenge in different contexts.

A standardized z-score (mean of 0 and standard deviation of 1) of perceived challenge was calculated and took into account individual differences while also allowing for comparison across individuals on a common scale. The z-scored perceived challenge also provides an advantage for exploring the emotional response in different contexts. The students' z-scores of challenge are compared across different settings and activities in these case studies. Additionally, the students' positive emotional states, which are measured by "happy, enjoy, excited, success and competitive," and negative emotional states, which are measured by "angry, stress, confused, give-up and anxious," are compared across different levels of the students' challenge levels using a different visualization. An average score of five emotional responses was used to represent the positive and negative states for the three cases. The five positive and negative emotional states were chosen based on earlier work (Hektner et al., 2007, pp: 110–123). These analyses are depicted through graphs to illustrate these varying states of challenge with the students' other positive and negative emotional states. These analyses give insight at the individual level; however, to understand aggregate relationships, we move to statistical models with the entire sample of ESM students from the field test and efficacy study.

From case studies to a statistical model

The ESM asks students questions that correlate perceived challenge experiences that may confound the relationship with other positive and negative psychological states. Therefore, it is important to understand the influence confounding variables may have on students perceived social and emotional

1 https://www.pacoapp.com/

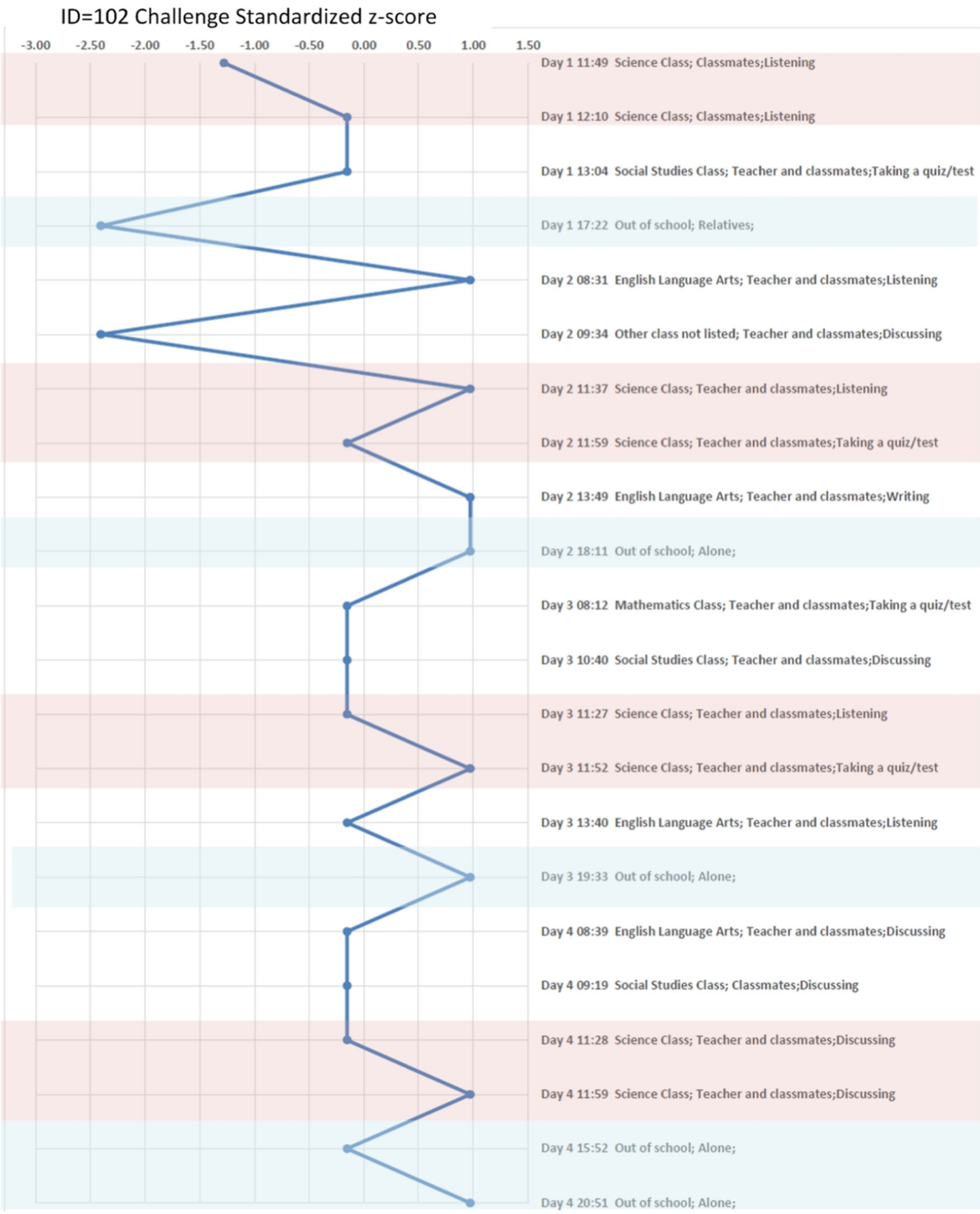

FIGURE 2
Dennis (Low-performing student) Experienced "Challenge" in situ across Context. Report Z-score Over 4 days. Pink color marks the moments in science classroom, and the light blue color marks the moments when a student is out of the school.

well-being. More specifically, in the case of perceived challenge, this new work has begun to examine the confounding effects of stress and anxiety on the relationship between challenge and two important negative psychological states, confusion and giving up, and one positive state of determination (Bradford and Bartz, 2022). First, the correlations for the variables were calculated to understand the relationship between challenge, stress, anxiety, confusion, giving up and determination.

Then, using a repeated measures HLM, the relationship between challenge and confusion, giving up, and determination was explored first without covariates and then with stress and anxiety as covariates. The following two equations were estimated.

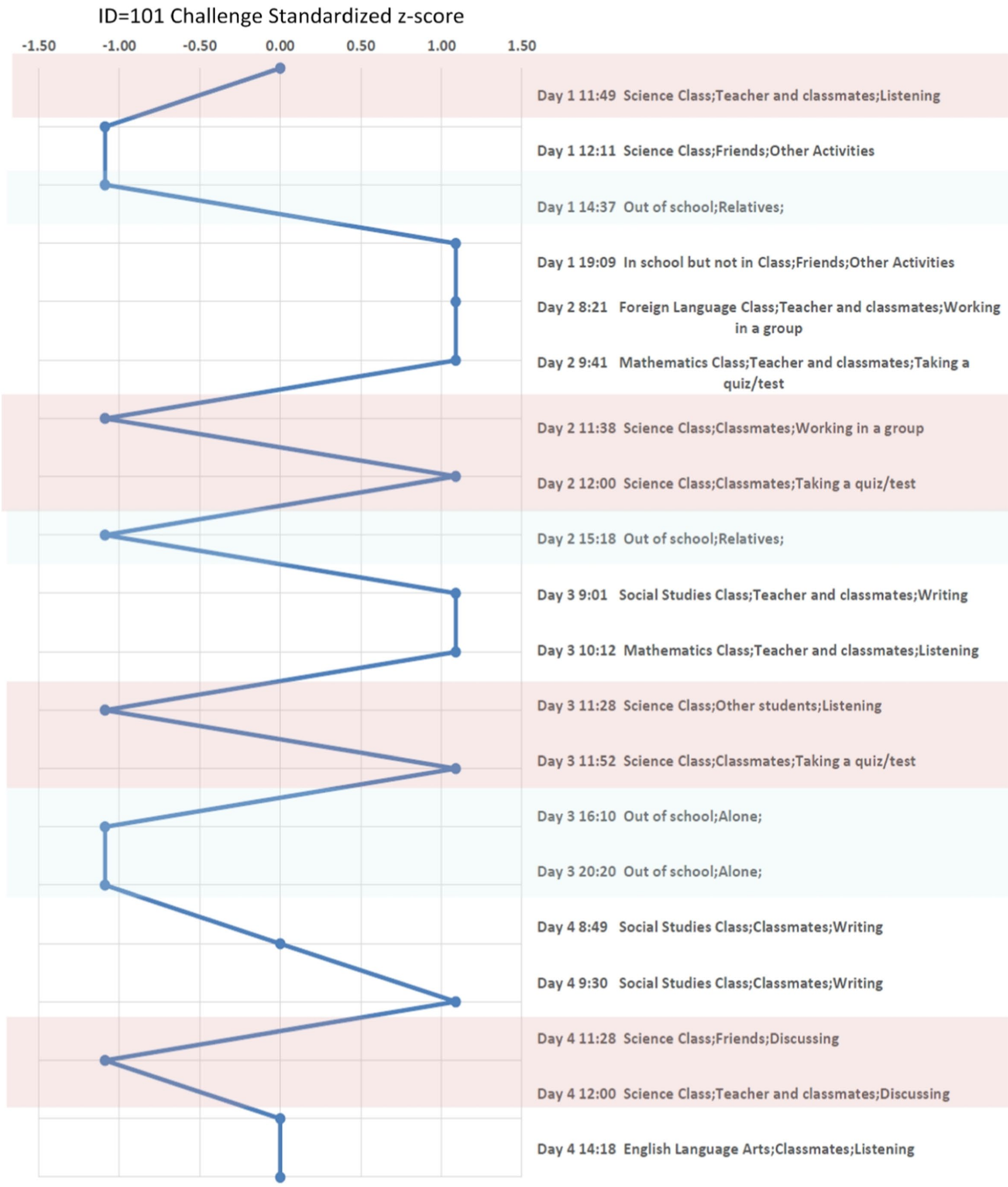

FIGURE 3
Megan (Average student) Experienced "Challenge" in situ across Context. Report Z-score Over 4 days. Pink color marks the moments in science classroom, and the light blue color marks the moments when a student is out of the school.

Model 1:

$$outcome_{ij} = \delta_{00} + \delta_{10} challenge + \nu_{0j} + \epsilon_{ij}$$

Model 2:

$$outcome_{ij} = \delta_{00} + \delta_{10} challenge + \delta_{20} stress + \delta_{30} anxiety + \nu_{0j} + \epsilon_{ij}$$

Where δ10 is the relationship between challenge and the outcome in the two models, v_0j is the student level random intercept and epsilon_ij is the beep level error term. The δ_{10} from both models 1 and 2 were compared using the Hausmann test to determine if the inclusion of stress and anxiety significantly changed the relationship between challenge and the outcome.

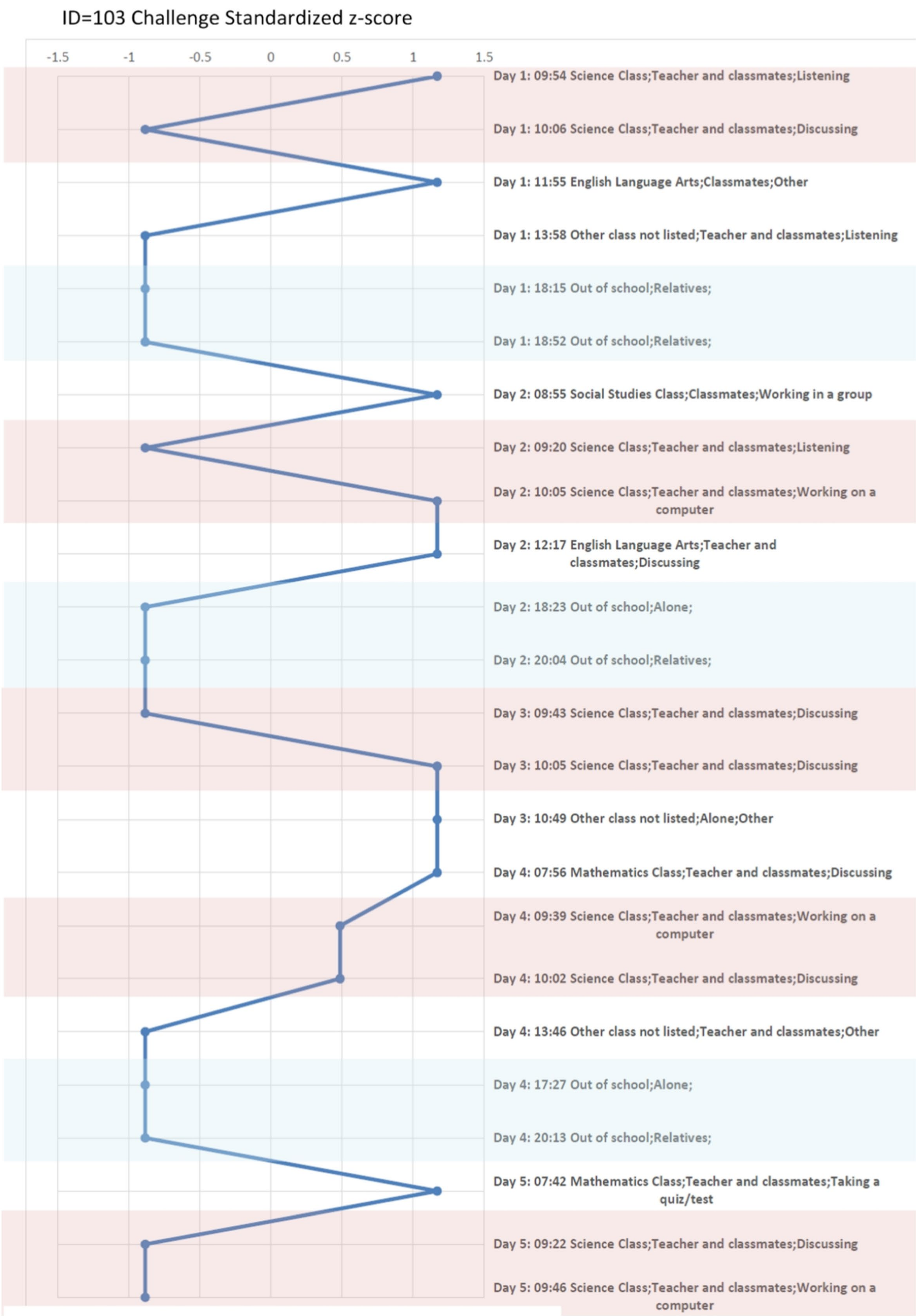

FIGURE 4
Collins (Above Average student) Experienced "Challenge" in situ across Context. Report Z-score Over 4 days. Pink color marks the moments in science classroom, and the light blue color marks the moments when a student is out of the school.

TABLE 5 Three student case study.

	School performance	Gender	Average Perceived challenge (individual) all context	Average Perceived challenge (individual) in science classroom	*n* of moment
Dennis	Low-performing student	Male	3.14	3.25	22
Megan	Average performance	Female	2.51	2.25	20
Collins	Above average	Male	2.29	2.3	24

Results

Three case studies

In Figure 2, we plot the z-score of Dennis' perceived challenge over 22 moments. We also plot the z-score of Megan' perceived challenge over 20 moments in Figure 3 and the z-score of Collins' perceived challenge over 24 moments in Figure 4. The three plots center in the middle line of the z-score as 0, which is each individual student's average challenge score. Dennis has a smaller range of perceived challenge than Megan and Collins, and he favors to report high perceived challenge among 4 days of the study. If we are interested in the context of science classroom, we can compare patterns when the three students are taking quiz in the same context. For example, we use the pink color to mark the moments in science classroom, and the light blue color mark the moments when a student is out of the school. Dennis and Megan perceived higher challenges, particularly when taking a quiz (z-score ranged from 0.5 to 1.0). Collins feels less challenged when taking a quiz but experiences a higher challenge in group discussion or when using a computer. We can conclude that Dennis and Megan objectively have higher perceived challenge than Collins when taking a quiz among the 4 days they were sampled.

Relative to the science classroom context, the out-of-school context (colored in light blue), Megan and Collins are less challenged especially when compared to Dennis. Overall, these three case studies show that the context of when students feel challenged can vary considerably by individuals and activities.

Recognizing the individual variability of experiencing challenge across contexts, the next question is whether emotional responses related to challenge differ by student. When challenged, is this experience more positive for Collins and Megan than Dennis, or do they all report similar feelings? Figures 5–7 show other positive and negative emotional states of these three students as well as their level of challenge during their science classes.

Among our three student cases, Dennis had fewer positive psychological states and reported more challenging tasks during his science class. His perception of high challenge (light blue bar) is less correlated with positive psychological states and more correlated with negative ones. Additionally, Dennis' psychological states fluctuated more than the two other students. These fluctuations are more apparent when experiencing positive psychological states (e.g., feeling successful, confident) than his negative psychological states of confusion.

Megan, on the other hand, had a declining trend of positive psychological states over the 4 days. When she experienced high challenge tasks in the science classroom, her negative psychological states increased. However, for Collins, the above average student, his positive psychological states were more correlated with higher challenge. Additionally, during moments of rising challenge, Collins experienced other positive emotions. Overall, the relationship between challenge and positive and negative psychological states seem to vary across students and days (Chen et al., 2022).

Statistical results from the entire sample

Among all students in the ESM sample, challenge was closely related to negative emotions for some students, while for others, it was closely related to positive emotions, and for others, there was no relationship. However, these results do not indicate how stress and anxiety might be influencing the relationship of challenge with other positive and negative emotions. Instead of focusing on all positive and negative emotions, a few key variables were explored more deeply: confusion, giving up, and determination, which were all positively correlated with challenge as seen in Table 6, which includes the entire sample of students, 1,412 students from the field test and efficacy study.

However, importantly, confusion and giving up were also positively related to stress and anxiety, while determination was not. Therefore, the question arose was whether stress and anxiety may be accounting for this positive relationship between challenge and confusion and challenge and giving up.

From the repeated measures HLM, the positive relationship between challenge and confusion and challenge and giving up significantly decreased in absolute value, when including stress and anxiety as covariates ($\delta \approx 0.1, p - value < 0.001$), while the relationship between challenge and determination remained relatively the same (Bradford and Bartz, 2022).

Discussion

The present findings extend previous research in at least two ways: First, these results provide a moment-level look at context differences in response to daily challenges in school, incorporating

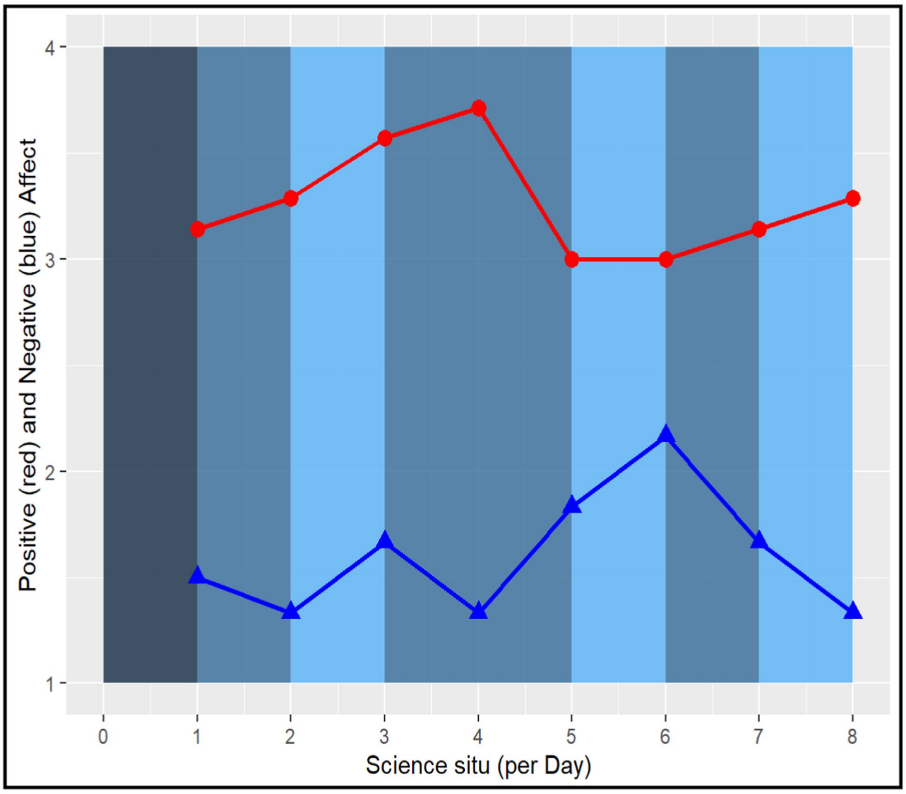

FIGURE 5
Dennis (Low-performing student) Experienced Challenge in Relation to Positive and Negative Psychological States. Report raw scores in positive and negative emotions. The dark blue color marks the lowest challenge moments (=1), whereas the light blue color marks the highest challenge moments (=4) in the science classroom.

both the intra-individual as well as the inter-personal level across study one and study two. Both offer insights for each other to complete the puzzle of challenging experiences in students' daily lives in school. These finding of significant context differences in intra-individual variability of experiencing challenge and other positive and negative states, to some degree, suggest that the relationship between perceived challenge, optimal learning moment, and psychological reactions is complex. Examining these relationships among these different emotions also offers that classroom learning, as we might have expected, is not a simple correlation with a specific experience but needs to be seen in context, over time, and in relationship to other events.

To further consider individual and contextual factors simultaneously, a designed statistical model like Simultaneous Equation Modeling (SEM) or Dynamic Structural Equation Modeling (DSEM) is essential to move this line of research forward. Second, complementing previous optimal learning moment literature on the states of the flow (Csikszentmihalyi and Csikszentmihalyi, 1988; Schneider et al., 2016), students' determination could be one psychological state that may keep students working on the challenging tasks in science. The results from adding stress and anxiety as covariates indicates that as one explores individual and contextual differences in their experiences, one should also consider the confounding effects that may occur when these individuals are experiencing many different emotional states at once. Additionally, these results may suggest that stress and anxiety may not be as important of an activation for challenge. These results offer some possibilities for discovering methods to increase students' optimal learning moments in science.

Limitations of the study

With respect to specific limitations of study 1, there are few '*in situ*' surveys for elementary level students for which we could compare our results. We have plans to use collected videos of classrooms to collaborate our findings which could increase the validity of these instruments. With respect to study 2, there may be other emerging technologies that could capture more changes in emotionality than the ESM, such as combining individual responses with video technology to capture facial, cognitive, and biomarkers, to which we could compare our results.

Overall, more studies are needed to use these techniques to build a corpus of work so that a comparison across studies can be examined to understand the reliability and validity of these techniques and their results. Despite our limitations of not having more in-depth analyses of personal and environment influences on social and emotional learning, our work provides another lens for understanding how levels of engagement and motivation are related to achievement, especially today when COVID's effects on

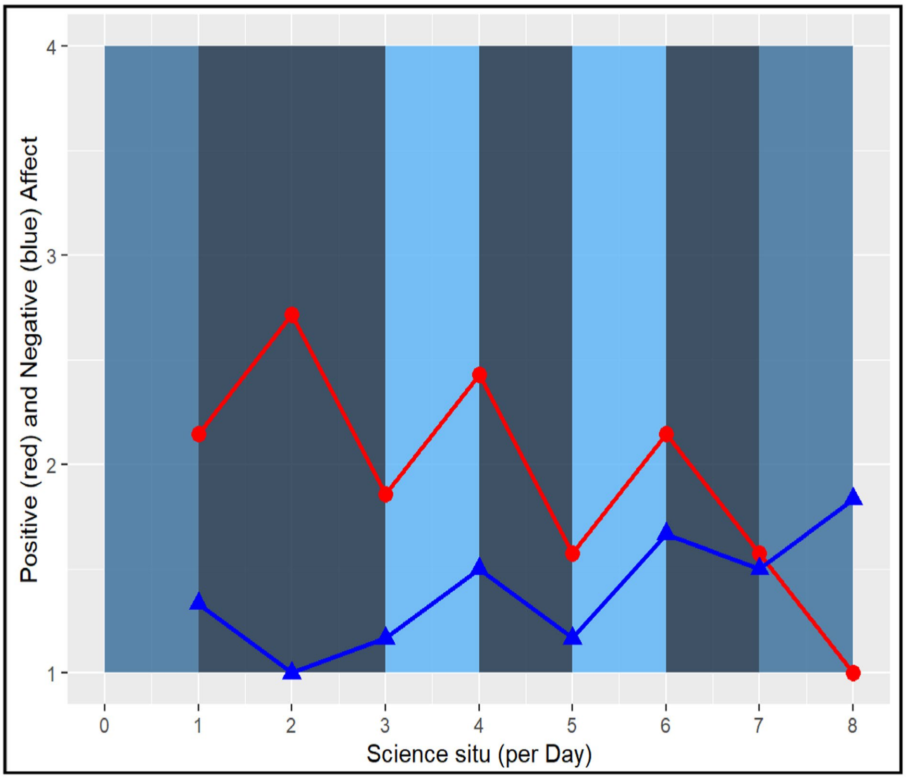

FIGURE 6
Megan (Average student) Experienced Challenge in Relation to Positive and Negative Psychological States. Report raw scores in positive and negative emotions. The dark blue color marks the lowest challenge moments (=1), whereas the light blue color marks the highest challenge moments (=4) in the science classroom.

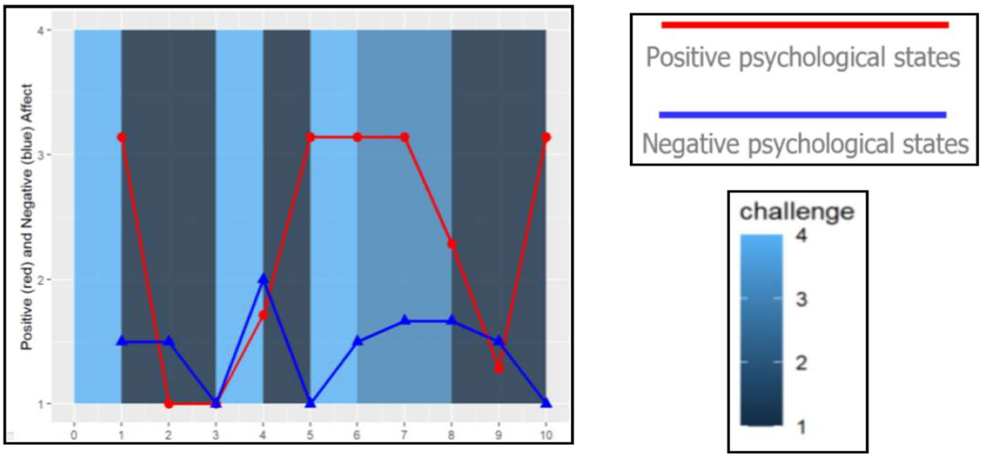

FIGURE 7
Collins (Above Average student) Experienced Challenge in Relation to Positive and Negative Psychological States (Low-performing student). Report raw scores in positive and negative emotions. The dark blue color marks the lowest challenge moments (=1), whereas the light blue color marks the highest challenge moments (=4) in the science classroom.

these important relationships need further exploration. Additionally, more studies are needed on emotionality in classrooms that take into account student cultures, family histories, race/ethnicity, and gender.

Implications for raising interest in STEM

Even though classrooms are busy fluid learning environments, results show it is possible to measure social and emotional

TABLE 6 Pairwise correlations of challenge, stress, anxiety, giving up, determination, and confusion.

	Stress	Anxious	Challenge	Give up	Determined	Confused
Stress	1.000					
Anxious	0.460	1.000				
Challenge	0.310	0.270	1.000			
Give up	0.430	0.310	0.380	1.000		
Determined	−0.028	0.093	0.220	−0.099	1.000	
Confused	0.54	0.400	0.420	0.490	−0.022	1.000

experiences, but they vary considerably by context. Some students find certain types of activities more interesting than others, and their skill levels vary meeting similar challenging problems with a diverse set of reactions from boredom and confusion to determination and sense of success and accomplishment. These variations by context indicate that isolating a specific measure of emotionality may overlook the factors at work that could be deterrents to motivation and persistence in STEM. It seems critical that researchers attempting to increase motivation for students to become engaged in learning experiences need to focus on the environment and emotions which operate at the same moments within the same context. And most importantly when considering engagement, recognizing that students vary in their skill levels, and this may be affecting the pursuit of learning new skills and attempting challenging STEM problems.

The greatest challenge for researchers who wish to transform STEM learning environments is determining the important types of social and emotionality constructs that make the most sense given the subject matter and experiences when students are expected to be engaged. This work has deliberately focused on science classrooms, where the underlying instructional and curricular activities are crafted in accordance with recent reports for transformative pedagogical practices. The toolkit of social and emotional measures being considered are those that seem the most reasonable given the goals of the lessons and the phenomena and problems to be solved. However, in trying to disentangle the behavioral, cognitive, and emotionality of engagement, considering interest, skill, and challenge are imperative, as well as other social and emotional factors that also occur when valuing teamwork and collaboration for having students learn and work with others and reach a place of ownership of ideas and products.

During the in-depth study of engagement (i.e., interest, skill, and challenge), several new factors related to learning have occurred. Interest, as others have also recognized, is critical; however, it must be constructed around ideas that the students find purposeful to their own lives. Memorizing the elements of the periodic table without knowing the purpose behind understanding the properties of the atom is a non- starter to a student. However, why we need to understand the relationship of certain elements to each other and their impact on chemical reactions experienced in everyday life can become more meaningful to a student.

Students have different skill levels and when choosing group experiences being attentive to the likely variation in the classroom is indispensable. The importance of bringing everyone into the problem-solving activity and making it a reasonable challenge for all students is likely to affect their personal as well as the groups' continued work on a project or problem. The idea here is not to construct activities that have the lowest level of skills but rather to offer various flexible routes to problem solving for all the students. Nonetheless, it is the case in PBL that there are certain disciplinary core ideas that are regarded as critical and that has to the starting point of the lessons. What students need is an awareness of their own confidence to face a challenge and how that can fit into the space of figuring out a phenomenon or solving a problem.

Moving students to learn something they do not know changes the nature of learning from memorization to using ideas. This type of learning poses another set of ideas, in that students are taking on something that they do not know but they could find out. This process exposes their vulnerabilities in of not knowing—for which they need to learn to be more comfortable with. This is particularly problematic for females especially in adolescence, where taking risks and exposing one's vulnerabilities is typically a positive aspect of the socialization process they encounter (Reniers et al., 2016). What is needed here is to underscore the value in taking intellectual risks in problem solving learning activities and the determination to continue working until a solution is found. Coming out of one's comfort zone intellectually particularly in science where discovery and new innovations are fundamental must be nurtured not just with content but the social and emotional factors that can inspire motivated students to solve. Understanding these relationships among these different emotions suggest that classroom learning as we might have expected is not a simple one to one correlation with a specific experience but need to be seen in context, over time and in relationship to other events. This underscores the difficulty and limitations of new curricular packages designed to measure and relate emotionality to achievement and certain positive behavioral actions.

Ethics statement

The studies involving human participants were reviewed and approved by Michigan State University IRB. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

BS: conception and design of work, drafting the article, review of data analysis and interpretation, and critical revision. I-CC: data analysis and interpretation, drafting the article, and critical revisions of the article. LB: data analysis and interpretation, drafting the article, and critical revisions of the article. KB: data analysis and interpretation, drafting the article, and critical revisions of the article. All authors contributed to the article and approved the submitted version.

Funding

This study is supported by the National Science Foundation (OISE-1545684; PIs Barbara Schneider and Joseph Krajcik); George Lucas Educational Foundation—Lucas Education Research (PI Joseph Krajcik); and the John A. Hannah Chair in the College of Education at Michigan State University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation and the George Lucas Educational Foundation.

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

Baines, A., DeBarger, A., De Vivo, K., and Warner, N. (2017). "Why is social and emotional learning essential to project-based learning?," *LER position paper 2* United States: George Lucas Educational Foundation.

Bartz, K., Miller, C., and Bateman, K. (2022). *Measuring Engagement in 3rd Grade Science Classes. Working Paper*. Michigan: Create for STEM.

Bradford, L., and Bartz, K. (2022). Intersections of challenge, stress and anxiety: When is the challenge too much? [Conference Presentation]. AERA Annual Conference 2022, San Diego, CA, United States.

Chen, I., Bradford, L., Bartz, K., and Krajcik, J. (2022). Exploring the stability and fluctuation of experiencing "challenges" in the high school science classroom [Conference Presentation]. AERA Annual Conference 2022, San Diego, CA, United States.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper Perennial.

Csikszentmihalyi, M., and Csikszentmihalyi, I. (1988). *Optimal Experience*. New York: Cambridge University Press.

Csikszentmihalyi, M., and Larson, R. (2014). "Validity and reliability of the experience-sampling method" in *Flow and the Foundations of Positive Psychology*. ed. Peele-Eady (Dordrecht: Springer), 35–54.

Csikszentmihalyi, M., and Schneider, B. (2000). *Becoming Adult: How Teenagers Prepare For the World of Work*. New York: Basic Books.

Cuddy, A. J. C., Fiske, S. T., and Glick, P. (2008). "Warmth and competence as universal dimensions of social perception: the stereotype content model and the BIAS map," in *Advances in experimental social psychology*. ed. M. P. Zanna, *Vol. 40.* (Oxford: Academic Press), 61–149.

Deaux, K., and La France, M. (1998). "Gender" in *Handbook of Social Psychology*. eds. D. T. Gilert, S. T. Fiske and G. Lindzey, vol. *Vol. 1. 4th Edn.* (New York: McGraw-Hill), 788–827.

Dorn, E., Hancock, B., Sarakatsannis, , and Viruleg, E. (2020). *COVID-19 and student learning in the United States: The hurt could last a lifetime.* Chicago, Illinois: McKinsey & Company Public sector Practice.

Durlak, J. A., Domitrovich, C. E., Weissberg, R. P., and Gullotta, T. P. (Eds.). (2015). *Handbook of social and emotional learning: Research and practice*. New York: Guilford Publications.

Evans, B. (2016). Paco—applying computational methods to scale qualitative methods. *EPIC.* 2016, 348–368.

Fortus, D., and Krajcik, J. (2012). *Curriculum coherence and learning progressions. Secon International Handbook of Science Education*. Berlin: Springer 783–798.

Fredricks, J., and McClolskey, W. (2012). "The measurement of student engagement: a comparative analysis of various methods and student self-report instruments," in *Handbook of Research on Student Engagement*. eds. S. Christenson, A. Reschly and C. Wylie (MA: Springer), 7–37.

Grabau, L. J., and Ma, X. (2017). Science engagement and science achievement in the context of science instruction: a multilevel analysis of US students and schools. *Int. J. Sci. Educ.* 39, 1045–1068. doi: 10.1080/09500693.2017.1313468

Hammerstein, S., König, C., Dreisörner, T., and Frey, A. (2021). Effects of COVID-19-related school closures on student achievement-a systematic review. *Front. Psychol.* 12:289. doi: 10.3389/fpsyg.2021.746289

Hektner, J., Schmidt, J., and Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oakes, CA: Sage Publications.

Immordino-Yang, M. (2015). *Emotions, learning and the brain: Exploring the educational implications of affective neuroscience*. New York: W.W. Norton.

Jagers, R. J., Rivas-Drake, D., and Borowski, T., (2018). Equity & social and emotional learning: a cultural analysis. Frameworks Briefs (November). Available at: https://casel.org/wp-content/uploads/2020/04/equity-and-SEL-.pdf

Kraft, M. (2020). Interpreting effect sizes of education interventions. *Educ. Res.* 49, 241–253. doi: 10.3102/0013189X20912798

Krajcik, J., and Schneider, B. (Eds.) (2021). *Science education through multiple literacies: Project-based learning in elementary school.* MA: Harvard Education Press.

Krajcik, J., Schneider, B., Miller, E., Chen, I., Bradford, L., Baker, Q., et al. (in press). Assessing the effect of the project-based learning on science learning in elementary schools. *Am. Educ. Res. J.*

Krajcik, J., Schneider, B., Miller, E., Chen, I., Bradford, L., Bartz, K., et al. (2021). Assessing the effect of the project-based learning on science learning in elementary schools. *Final Technical Report to Lucas Education Research*

Krapp, A., and Prenzel, M. (2011). Research on interest in science: theories, methods, and findings. *Int. J. Sci. Educ.* 33, 27–50. doi: 10.1080/09500693.2010.518645

Lavonen, J., Juuti, K., Uitto, A., Meisalo, V., and Byman, R. (2005). *Attractiveness of science education in the Finnish comprehensive school*. Helsinki, Finland: Technology Industries of Finland.

Lee, C., Nasir, N., Pea, R., and de Royston, M. (2019). Introduction Reconceptualizing learning: a critical ask for knowledge building and technology. In N. S. Nasir, C. D. Lee, R. Peas and M. M. de Royston (Eds.) *Handbook of the cultural foundations of learning.* (United Kingdom: Rutledge), 17–35.

Lerner, R. M., and Steinberg, L. (Eds.). (2009).*Handbook of adolescent Pshycology*. (*3rd ed.*). United States: John Wiley & Sons Inc.

Miller, E. C., and Krajcik, J. S. (2019). Promoting deep learning through project-based learning: a design problem. *Disciplinary and Interdisciplinary Science Education Research* 1, 1–10. doi: 10.1186/s43031-019-0009-6

Moeller, J., Spicer, J., Salmela-Aro, K., and Schneider, B. (2017). "Advances in the research on situation-specific and contextual aspects of student engagement," in *Pathways to adulthood, educational opportunities, motivation, and attainment in times of social change*. eds. I. Schoon and R. K. Silbereisen (London: UCL IOE Press), 119–136.

National Academies of Sciences. (2021). *Science and engineering in preschool through elementary grades: The brilliance of children and strengths of educator's*. United States: National Academies Press.

National Academis of Sciences. (2018). *How People Learn II: Learners, Context, and Cultures*. Washington, DC: National Academies Press.

National Assessment of Educational Progress (2021). *Results From the 2019 Science Assessment*. U.S. Department of Education and Institute of Education Sciences.

National Research Council. (1999). *How People Learn: Brain, Mind, Experience, and School*. United States: National Academies Press.

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. United States: National Academies Press.

Organisation for Economic Co-operation and Development. (2020). Science performance (PISA)—Indicator. Available at: https://data.oecd.org/pisa/science-performance-pisa.htm

Peele-Eady, T. B., and Moje, E. B. (2020). "Communities as contexts for learning," in *Handbook of the Cultural Foundations of Learning*. eds. N. S. Nasir, C. D. Lee, R. Peas and M. M. de Royston (United Kingdom: Routledge), 230–246.

Reniers, R. L. E. P., Murphy, L., Lin, A., Bartolomé, S. P., and Wood, S. J. (2016). Risk perception and risk-taking behaviour during adolescence: the influence of personality and gender. *PLoS One* 11:e0153842. doi: 10.1371/journal.pone.0153842

Renninger, K. A., and Hidi, S. E. (2020). To level the playing field, develop interest. *Policy Insights Behav. Brain Sci.* 7, 10–18. doi: 10.1177/2372732219864705

Rosado-May, F. J., Urrieta, L. Jr., Dayton, A., and Rogoff, B. (2020). "Innovation as a key feature of indigenous ways of learning: individuals and communities generating knowledge 1" in *Handbook of the cultural foundations of learning.* eds N. S. Nasir, C. D. Lee, R. Peas and M. M. de Royston (United Kingdom: Routledge), 79–96.

Salmela-Aro, K., Moeller, J., Schneider, B., Spicer, J., and Lavonen, J. (2016). Integrating the light and dark sides of student engagement with person-oriented and situation-specific approaches. *Learn. Instr.* 43, 61–70. doi: 10.1016/j.learninstruc.2016.01.001

Schneider, B., Krajcik, J., Lavonen, J., and Salmela-Aro, K. (2020). *Learning science: The value of crafting engagement in science environments*. New Haven: Yale University Press. Also published in Chinese by, Educational Science Publishing House Limited, 2021.

Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Broda, M., Spicer, J., et al. (2016). Investigating optimal learning moments in U.S. and Finnish science classes. *J. Res. Sci. Teach.* 53, 400–421. doi: 10.1002/tea.21306

Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Klager, C., Bradford, L., et al. (2022). Improving science achievement—is it possible? *Educ. Res.* 51, 109–121. doi: 10.3102/0013189X211067742

Shernoff, D., Csikszentmihalyi, M., Schneider, B., and Shernoff, E. (2003). Student engagement in high school classrooms from the perspective of flow theory. *Sch. Psychol. Q.* 18, 158–176. doi: 10.1521/scpq.18.2.158.21860

Shumow, L., and Schmidt, J. A. (2014). *Enhancing Adolescents' Motivation for Science: Research-Based Strategies for Teaching Male and Female Students*. Thousand Oaks, CA: Corwin, A Sage Company.

TYPE Community Case Study
PUBLISHED 19 December 2022
DOI 10.3389/fpsyg.2022.924943

OPEN ACCESS

EDITED BY
Katja Upadyaya,
University of Helsinki,
Finland

REVIEWED BY
Joseph Roche,
Trinity College Dublin,
Ireland
Duhita Mahatmya,
The University of Iowa, United States

*CORRESPONDENCE
Tamara A. Franz-Odendaal
Tamara.franz-odendaal@msvu.ca

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Gender, Sex, and Sexualities, a section of the journal Frontiers in Psychology

RECEIVED 20 April 2022
ACCEPTED 09 August 2022
PUBLISHED 19 December 2022

CITATION
Franz-Odendaal TA and Marchand S (2022) Girls Get WISE—A programming model for engaging girls+ in STEM.
Front. Psychol. 13:924943.
doi: 10.3389/fpsyg.2022.924943

Girls Get WISE—A programming model for engaging girls⁺ in STEM

Tamara A. Franz-Odendaal[1*†] and Sally Marchand[2†]

[1]Department of Biology, Mount Saint Vincent University, Halifax, NS, Canada, [2]WISEatlantic, Mount Saint Vincent University, Halifax, NS, Canada

The majority of STEM disciplines in Canada are male-dominated and there is a significant lack of programming available to girls. The Girls Get WISE program is a university-based program that is funded by the federal government, the university, and corporate sponsorship. This program is delivered in person by educational professionals, science students, and past participants. By engaging girls in hands-on interactive STEM activities in a safe and fun space, this program provides an opportunity for young women to showcase their talents and excitement for science-based topics. The features of this program and its evaluation over a 10 year period are described here.

KEYWORDS

outreach, middle school, science—general, STEM program, girls, science

Introduction

The Girls Get WISE programs began in 2011 at a university in Halifax, Nova Scotia, Canada. The city of Halifax has a population of about 450,000 people and is located on the East coast of Canada. The university that delivered the Girls Get WISE program has a long-standing reputation for the advancement of women. At the time when this outreach program was started, there was very limited STEM programming in the region, and only one other non-profit organization running STEM programming exclusively for girls, and that program was primarily focused on getting girls interested in Trades and Technology careers. The limited number of other STEM programs in the region are mixed gender and it was clear that, in most, the percentage of girls participating was 10% or less. The university itself had no science outreach programs.

Context

Statistically in Canada, the percentage of women working in STEM careers in Canada is low and is in stark contrast to the almost equal gender split in the labor workforce (Pereault et al., 2018). This percentage is not surprising given that male Science, Technology, Engineering, and Mathematics (STEM) graduates in Canada are more likely than female STEM graduates to work in STEM (Frank, 2019). In the 2016 Statistics Canada census, a third of men (37.5%) with a Bachelor's degree had studied in STEM whereas only 15.3% of women with a Bachelor' degree studied in STEM. Within STEM disciplines there are large differences in participation. For example, while over half of the men (52.2%) with a STEM Bachelor's degree studied engineering or engineering technology, only 25.4% of women selected these fields. The percentage of professional women engineers across the country has remained steady at around 17%–20% for several years (Engineers Canada, 2020). Over the last 10 years, in all STEM disciplines, only small shifts in the percentages of women's participation have occurred. Among university students in Canada, the percentages of women across disciplines similarly vary dramatically. Again, within Biology, female students account for about 62% of students, while only 30% in Physics (Pereault et al., 2018). University and government institutions are only starting to collect and report intersectional data to explore the participation of racialized women, indigenous women, and disabled women in these fields.

The career progression to achieving a STEM career starts with elementary and secondary school. In Canada, most public schools are mixed gender with the majority of students attending public schools. The Science curriculum across the country is varied with each province regulating and determining course content. Science is taught as a single subject from elementary through to junior high school in Nova Scotia, Canada. In high school (grades 10–12), students can select Biology, Chemistry, and/or Physics. Mathematics is required in each year of schooling from primary through to grade 12, however the amount of math (one to three courses) varies depending on the student's interests. Science teachers have limited time and resources to conduct hands-on activities in the classroom, especially in junior high (Grades 7–9) and high school (Grades 10–12). Grade 9 of high school is when students begin to select courses for the following school year and this is when students are either selecting or not selecting science-based courses. The choices made in Grade 9 for Grade 10, dictate the science options available to them in Grade 11 and Grade 12. All Science degree programs in Canada require Science pre-requisite courses from high school. Therefore a student without science subjects in Grade 12, has to spend an extra year or two obtaining these course credits before they are accepted into university science programs. Furthermore, university education is very expensive in Canada and is not affordable to many students. Thus, girls who pursue STEM-based programs at university are often those who (i) are encouraged by science and mathematics teachers and/or parents; (ii) are from affluent families who can afford university tuition and tutors (if needed) during high school; and (iii) who are exposed to STEM careers through family members. Programming, such as the Girl Get WISE events, is needed because the STEM stereotype is heavily white male-dominated in Canada and is reinforced by the branding of toys, clothing for children, and the voices/images portrayed in the media (e.g., Steinke, 2017). Thus, while many girls are interested in STEM subjects, they cannot see themselves in these careers because of the strong stereotype that still exists. Girls also have limited opportunities to learn that there are other girls interested in these subjects and that you do not have to be the top student in their science and mathematics class to have a career in STEM. Previous research has shown that girls tend to be most interested in careers that help society (e.g., Heaverlo et al., 2013; Franz-Odendaal et al., 2020; Aviolo et al., 2022). STEM professionals, the media, and STEM organizations need to do a better job at portraying their fields as helping society and to shift the additional stereotype of this as lonely work.

Materials and methods

Programming overview

The two signature Girls Get WISE events described below are a one-day Girls Get WISE Science Retreat and Junior and Senior Girls Get WISE Science Summer Camps. These events are open to individuals who identify as girls (i.e., girls+). Both of these events feature hands-on STEM activities as well as an hour-long Role Model session. The hands-on sessions are always very interactive and ensure the girls are learning some aspect of STEM in a novel and fun manner. The hands-on activities are either developed in-house by a team of science-trained professionals or by others (i.e., science organizations or graduate students, etc.). In the latter case, all activities are thoroughly reviewed prior to accepting them into the program content (see Discussion). The Role Model session works round-robin style; five to six women working in different male-dominated STEM fields are invited to chat informally with small groups of participants about their careers for about 8 min, and then the girls switch to the next role model. Included with the recruitment of potential role models is a "Role Model Guide" guide document, which details how the session will run, as well as topics they are encouraged to cover such as their career pathways, what motivated them to pursue their careers, what their typical day is like, any setbacks they had along the way, etc. The girls are encouraged to ask questions, and question prompt cards are provided to them at the tables. Participants are informed that they can always reach out to the Program Coordinator to connect with specific role models if they have further questions after the session. It is important for girls to see people like themselves working in male-dominated fields to show them that these careers are possible.

The one-day Girls Get WISE Science Retreat brings together 50–60 girls+ in grades 7–10 (ages ~12–16 years) together at Mount Saint Vincent University (Nova Scotia, Canada) to participate in two hands-on STEM sessions and a Role Model session. The day starts with a team ice-breaker activity, which is typically an engineering design challenge or STEM trivia, and then the girls are split into two random groups to participate in their first hands-on activity of the day which is typically an hour in length. There is a break for an hour lunch (that is provided), then they move to the second hands-on activity of the day, which is then followed by the hour-long Role Model session. At least one of the hands-on activities takes place in a laboratory setting each year, and the other session could be coding or Engineering related. The day ends with some reflection, an evaluation, and prize draws.

From 2012 to 2016, one Girls Get WISE Science Summer Camp took place each year for 24 girls+ that were 12–14 years in age. In 2017, a second Science summer camp for older girls, 15 and 16 years in age, was added. Both camps are day camps; however, the Junior camp runs for a full 5 days, while the Senior camp runs 4 days a week and at reduced hours per day. The difference in length is because the younger girls are more likely to depend on parents or caregivers for transport to the camp, while older girls are more able to take public transport. Similar to the Science Retreats (described above), both of the camps feature an hour-long Role Model Session, as well as different hands-on STEM activities, some of which take place in the laboratory, while others take place outdoors. Figure 1 shows an example of one of the programs for a science camp.

The major difference between the Junior and Senior camps is that the Junior camps introduce the participants to the field of Developmental Biology or Embryology through the study of zebrafish embryos. The reason for this is that the program chair is a researcher studying zebrafish development and as such these fish areas are available on campus. Zebrafish (*Danio rerio*) are wonderful organisms to showcase to students as they are easily obtainable (e.g., from pet stores), they breed readily and produce hundreds of eggs per clutch, their development is external to the mother, and their embryos are transparent (Wilk et al., 2018). The camp participants are first taught the traits of the zebrafish that make them an ideal model organism, the environmental parameters needed for their survival, and how to handle the embryos. Participants are then given several embryos from a clutch of embryos to study for the duration of the week. This teaches the participants the importance of observation and note-taking in science. The girls are provided lab-books to take notes and to draw the embryos on a daily basis. It also allows the girls to become comfortable in a laboratory setting within a university. Participants in the Junior camp spend about 2 h a day in the lab observing and caring for their

Schedule

Monday

9:00 - 9:30	Introduction
9:30 -10:45	Water of the World (Engineering/Math)
10:45-11:00	**BREAK**
11:00-12:00	*In the Lab* (Biology)
12:00- 1:00	**LUNCH**
1:00 - 2:30	Art Session (Biology)
2:30 - 3:00	Introduction to Zebrafish Development (Biology)
3:00 - 3:30	**BREAK**
3:30 - 4:00	The Scientific Approach + Breeding (Biology)
4:00 - 4:30	Data Collection Part I (Biology)

Tuesday

9:00 - 9:30	*In the Lab: collecting eggs* (Biology)
9:30-10:00	*In the Lab* (Biology)
10:00 -10:15	**BREAK**
10:15 -11:00	*In the Lab* (Biology)
11:00 -12:00	Leadership Skills
12:00 - 1:00	**LUNCH**
1:00 - 2:30	Egg Drop (Physics/Math)
2:30- 2:45	**BREAK**
2:45- 3:30	Chemistry Session
3:30- 4:30	*In the Lab* (Biology)

Wednesday

9:00 - 10:30	*In the Lab* + *Photography of Zebrafish* (Biology)
10:30 - 10:45	**BREAK**
11:00 - 12:00	App Design (Technology)
12:00 - 1:00	**LUNCH**
1:00 - 2:00	App Design (Technology)
2:00 - 3:00	*In the Lab* (Biology)
3:00 - 3:30	**BREAK** + Ice Cream Social
3:30 - 4:00	Career Survey (Research)
4:00 - 4:30	*In the Lab* (Biology)

Thursday

9:00 - 10:00	*In the Lab* (Biology)
10:00 - 10:15	**BREAK**
10:15 - 11:30	Entrepreneurship (CEED)
11:30 - 12:00	Career Aspirations (Guidance)
12:00 - 12:45	**LUNCH**
12:45 - 1:45	Role Model Session (Networking/Guidance)
2:00 - 2:30	Hatching Party
2:30 - 3:30	Explore Webinar (Engineering)
3:30 - 3:45	**BREAK**
3:45 - 4:30	*In the Lab* (Biology)

Friday

9:00 -10:00	*In the Lab* (Biology)
10:00 -10:15	**BREAK**
10:15- 11:45	Data Collection Part II (Biology/Math/Technology)
11:45 - 12:45	**LUNCH**
12:45 - 1:45	Steam Plant Tour (Engineering/Physics)
1:45 - 3:30	Camp Conclusion and Preparation for Open Lab
3:30 - 4:30	***Open Lab with Family***

FIGURE 1
Example of a Junior Girls Get WISE Science Summer Camp program of events.

zebrafish embryos. After this, the hatched larvae are returned to the researcher's fish facility.

Another aspect of the Junior Girls Get WISE Science Camp is on the final day of camp, participants are encouraged to invite their families, friends, and community members to campus for an "open house." Participants display STEM activities that they completed throughout the week and then showcase these to the guests. Parents are one of the key influencers on youth's decision to pursue a STEM career or not (e.g., Dasgupta and Stout, 2014; Franz-Odendaal et al., 2016), so it is critical that parental figures are incorporated into the program. This is done by showcasing the work and projects the students have completed in the camp to parents and grandparents (and other family members and guardians) on the last day of camp. These individuals thoroughly enjoy seeing the projects that they have heard about all week from their daughters.

Other than the zebrafish component, the Junior and Senior camps have similar activities, although each is geared to the appropriate age group. On the first day of both camps there is an icebreaker activity, then a group engineering design challenge. This is done because these activities promote teamwork and critical thinking, and provide an informal opportunity to meet the other girls+ in the camp. We strive to create a program that has a balance of different Chemistry, Biology, Physics, and Technology activities throughout the week. This includes a balance between indoor and outdoor activities. The university campus is fortunate to have a beehive and beekeeper, as well as a community garden, so activities involving those resources are often developed. Students are kept very busy throughout the camp and the level of excitement increases on a daily basis.

Pre-planning and delivery

Program development for the Science Retreats and Camps is led by the Program Coordinator with assistance from part-time staff, which in most cases are past participants of the programs. The Program Coordinator has degrees in both Science and Secondary Education. Past participants can apply to volunteer at a camp the following year. If they are interested in assisting thereafter then they are paid a standard student rate. Programming development typically begins 4 to 5 months in advance for Retreats and Camps. After selecting a date and booking the required spaces on campus. The promotional materials, which are used on social media, are prepared. These are also sent directly to Junior and Senior high schools in the area, as well as past participants. Registration is online and is first-come, first-served. It is often the case that a waiting list is required due to the high demand.

The cost for these events are very low fees: C$10 per participant for the Retreat and C$100 for the Camps. Advertisements also state that if cost is a barrier, then guardians can contact the Program Coordinator directly for a fee waiver. The aim is to make these experiences available to those of all socioeconomic backgrounds as research in both the United States (Afterschool Alliance, 2021) and Canada (Duodu et al., 2017) indicates that the cost of programming is one of the barriers youth from low-socioeconomic backgrounds face when it comes to attending after-school STEM programming. The registration fee does not cover all the costs to run the program and was implemented as a commitment to attend rather than as cost recovery. The registration fees are used to purchase materials for the activities, prizes for the participants, and thank you gifts for role models and volunteers.

The program continually recruits role models for the Role Model Sessions through a sign-up form on the program's website, word-of-mouth, social media posts, local universities, partnerships with industry, and not-for-profit organizations. If someone is interested in participating as a role model for the program we add them to the database of role models. When it comes time to plan a Girls Get WISE event, role models are selected from this database. We prioritize role models that have not participated recently (this reduces role model fatigue), that work in different STEM fields (to ensure a diversity of careers and career paths are showcased), and that come from diverse backgrounds (e.g., racialized women). Invitations to participate are sent to five or six role models. Sometimes, the timing of the event does not align with their schedule, for example, and in these cases, our role model database is consulted once again. We found that it was extremely useful to have a database of role models we could call upon at any time since a number of repeat participants attend the events, and therefore, both the program content and the selection of role models need to be different in consecutive years.

Delivery of the majority of the hands-on sessions is carried out by the Program Coordinator, part-time students and/or volunteers. Often experts are invited to discuss their research with participants and to then lead related hands-on activities, this typically occurs if the team is not familiar with a particular STEM topic. We also invite other STEM-related non-profit groups to host activities, such as the local Science Centre or Canada Learning Code. As noted earlier, it is important that the Program team reviews the activities to ensure that they are age appropriate.

Results

Program impacts

At all of these Retreats and Camps, evaluations are handed out at the end of the programming on the final day. These evaluations are primarily used to determine which sessions were well received and which need improvement. We also ask some program impact questions, although we acknowledge that these are leading questions. The past participants (from the last 10 years of programming) that completed program evaluations described below range in age between 12 and 16 years of age, they all identify as young women (girls+), and they come from varied socioeconomic backgrounds. The majority of respondents live in urban areas (roughly 85%). In total, 678 girls have participated in this program over the last 10 years and completed evaluation forms at the end of the events. Since these forms are handed out in the last few hours of the event, the response rate is high (95%). The data presented below is aggregated data from the last 10 years.

One of the questions participants are asked is "Did this science camp (or retreat) meet your expectations?" Participants could select from the following options: Exceeded, Yes, or No. From responses, 96% of participants indicated that the event met or exceeded their expectations (ANOVA: $F = 22.11$; $p < 0.005$). To better understand if the Girls Get WISE programming has a lasting impact on participants, participants are asked: "Did attending this event inspire you to continue with science at school?" Participants could select from the following options: Yes, No, Maybe, Unsure. 93% of participants indicated that the event did inspire them to continue with science at school (ANOVA: $F = 43.97$; $p < 0.005$). Another question is: "How did attending this event affect your interest in science and engineering?" of which participants could choose the following options for their response: More Interest, Same Interest, Less Interest. Results from this question were as follows: 60.7% of participants indicated that the event increased their interest, while 36.9% indicated that their interest stayed the same (ANOVA: $F = 5.12$, $p < 0.05$). Based on these responses, the many emails we receive after events and participant comments, it is evident that this programming style is highly successful; the longer girls stay on this science path the more exposure they would have to science and engineering as a career.

Also part of these post-event evaluations, participants are asked to rate each session overall on a scale of 1 to 5 with 1-disliked to 5-excellent. The role model session is consistently rated as one of the top sessions at these events by participants, getting an average rating of 4.3 out of 5. When asked the open-ended question "Do you have any additional comments or suggestions?" many participants mention the role model session in particular. Some comments received include: "*I would love more time to talk with the role models. They were so Inspiring*.," "*Role models were very fun and loved to hear them*," and "*The role model session helped to round out some questions I had about university*."

It is clear from the event evaluations that the STEM activities that were rated the highest are also those that involve using equipment few would have access to in schools, and those that are very hands-on. A few activities that were rated 4.5 or higher out of 5 were: studying the zebrafish in the lab, making soap with a local soap maker, working with planaria, making bath bombs, microscope scavenger hunt, and budgeting for life. Three of these activities, working with zebrafish, planaria, and the scavenger hunt, all involve the use of microscopes. Planaria are a type of flatworm that are able to regenerate its tissues, in this activity participants are asked to predict what will happen when certain parts of the work are cut away. They then perform the cuts and observe what happens over several days while caring for the planaria. The microscopes scavenger hunt has participants work in pairs and use written clues that match with prepared microscopes slides to make an educated guess as to what is on the slide. At the university, working with invertebrates such as planaria or with zebrafish embryos does not require animal ethics approval.

The one activity that differs from the others in the list above is "Budgeting for life." This activity was developed based on the game of life where participants are randomly given a job with an average salary and are asked to create a monthly budget using a template in Excel. This activity was designed so that participants were able to see the financial benefits that the majority of STEM careers can have over non-STEM career fields. Empowering the girls to be independent and financially secure women is an added benefit of a STEM career. A discussion about this activity which includes stressing that their passions and interests should be the primary determinant of their future career is conducted after the activity.

The Girls Get WISE events retain a good number of participants. On average, approximately 28% of participants attend more than one of the Girls Get WISE events. It is important to note that with a narrow age range for the events, many girls age out of the program within a year or two.

Discussion

Since the introduction of the Girls Get WISE programs in 2011, there has been an increase in STEM programming focused on engaging girls in Halifax, across the province, and in neighboring provinces. It is exciting to see this expansion in programs, especially to more rural parts of the region where it is difficult for the programs to reach on a consistent basis. A takeaway from this increase in all-girl programming is that parents are looking for these opportunities for their daughters and that funders are seeing the benefits and want to keep the momentum going.

Throughout the 10 years of programming, we have strived to utilize the feedback gathered from participant evaluations to improve on these program offerings. For instance, if a particular hands-on STEM session is rated less than three out of five on a five-point Likert scale, then the activity is reviewed and discussed to determine ways to improve that activity with the available knowledge and resources. If it could not substantially be improved, then the activity would no longer be offered. An example is that, after receiving suggestions from several camp participants that the number of hours spent in the laboratory studying zebrafish development was a little too long, we decreased the time for these sessions for all future camps. To ensure the success of this program, it is essential that activities are evaluated on a regular basis.

In the first few years of this program, the majority of the hands-on sessions were developed by the program team, but over time, external subject matter experts from the area were recruited to develop and run some of the activities. This was done because (i) it is time consuming to develop new activities, and (ii) the activities the team wanted to run were not in their field of expertise. This approach is two-fold, participants are able to meet additional STEM experts and the activities are often more tailored on a particular topic. This approach comes with other challenges, however. First, just because someone is an expert in a particular field or topic, does not mean that they have the experience in delivering an activity to youth based on that knowledge. To address this challenge, an activity proposal sheet was developed in which prospective facilitators were asked to complete. This form asks them to detail their activity and provide all relevant materials that would go along with it. The program team then works with the facilitators to tweak the content if needed. This proved to be an effective way to develop quality and age-appropriate content for the program.

Some lessons that were learned along the way in delivering all-girls STEM programming over the last decade are that it is important to create activities that are as hands-on and as interactive as possible, use the knowledge of local STEM experts to develop unique programming, and include them in the delivery if possible, including female role models as part of the programming is very impactful, and always strive to improve upon programming using feedback from participants (Figure 2). Specific to the Role Model Session, the ideal time for each "round" with a role model is seven to 8 min, over that amount is too long and the participants start to fidget, and under that time does not provide enough time for meaningful connection. The retreat

What Works for STEM Experiential Learning

- ✓ Hands-On and Interactive
- ✓ Involving Local STEM Knowledge Experts
- ✓ Female Role Models in STEM Fields
- ✓ Sharing Resources

FIGURE 2
A list of "What Works" for planning and delivering STEM experiential learning events.

format is quite popular and our team often serves as a resource to other groups wishing to run similar all-day events. Sharing resources is key to expanding reach.

While the school system curricula have remained largely unchanged over the last 10 years, there is a need to depend on other organizations to encourage the pursuit of STEM. Past research has shown that girls are influenced by their parents and guardians, teachers, and their peers (e.g., Franz-Odendaal et al., 2020). When speaking to the participants, peer influence appears to be the main driver in their decision to participate or not, in a STEM camp.

To build on these programs in the future, we would recommend the creation of an ambassador's program, where past participants become program ambassadors and through video and other social media content encourage their peers to participate in the program. Another program that would expand on the Girls Get WISE model would be to create a Girls Get WISE Leadership program, where past participants can get leadership training and then begin to volunteer with the programs. This helps to build confidence in the girls and would help to ensure a steady stream of ready and eager volunteers for STEM outreach programming.

The longer-term impact of these programs on the STEM workforce or student enrollment into STEM programs is unknown at this time. However, some insight can be gained from some of these unsolicited emails that were received: In the words of one of our participants "*I just wanted to touch base to let you know that the camps had an impact on my academic direction. I have just begun my first year in a Bachelor of Science degree in Biology at .. I wanted to thank-you for the opportunities we had during the WISE camps to better understand the different aspects of science and its various influences. It opened my eyes to different career possibilities. I am enjoying my first month at … and wanted to reach out to let you know these wonderful camps do make a difference.*"

And another email from a father of a past participant: "*My two daughters did the WISE Atlantic program a few years ago. My oldest has just been accepted to … University for Biology/Medical Sciences with a scholarship. I give great credit to your program for inspiring her to study and enjoy science. Many thanks for your program and the people who work to make it possible.*"

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.

Author contributions

TF-O and SM designed and delivered the programming and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Funding was received from the NSERC Chair for Women in Science and Engineering (Atlantic region). This program is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), Mount Saint Vincent University, Lockheed Martin Canada.

Acknowledgments

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC), Mount Saint Vincent University, Lockheed Martin Canada whose financial contributions made Girls Get WISE programming possible. We would also like to acknowledge the contributions of all the team members, partner organizations, volunteers, and role models who contributed their time and talents to making these events so successful.

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

Afterschool Alliance. (2021). STEM learning in afterschool on the rise, But barriers and inequities exist. Available at: www.afterschoolalliance.org/documents/AA3PM/AA3PM-STEM-Report-2021.pdf

Aviolo, B., Chávez, J., and Román, C. V.-S. (2022). Factors that affect the science and technology career choice decision of women high school students: a phenomenological study in Peru. *J. Women Minorities Sci. Eng.* 28, 27–55. doi: 10.1615/JWomenMinorScienEng.2021034228

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

Duodu, E., Noble, J., Yusuf, Y., Garay, C., and Bean, C. (2017). Understanding the delivery of a Canadian-based after-school STEM program: a case study. *IJ STEM Ed* 4:20. doi: 10.1186/s40594-017-0083-2

Engineers Canada. (2020). Percentage of newly licensed engineers who are women 30 by 30 data by C. Polyzou. Available at: https://engineerscanada.ca/diversity/women-in-engineering/30-by-30 (Accessed 2 March 2022).

Frank, K. (2019). "Analytical Studies Branch Research Paper Series," in *A Gender Analysis of the Occupational Pathways of STEM Graduates in Canada.* (Ottawa, Canada: Statistics Canada).

Franz-Odendaal, T. A., Blotnicky, K., French, F., and Joy, P. (2016). Experiences and perceptions of STEM. Subjects and careers and engagement in STEM activities Among middle school students. *Can. J. Sci. Math. Technol. Educ.* 16, 153–168. doi: 10.1080/14926156.2016.1166291

Franz-Odendaal, T. A., Blotnicky, K., and Joy, P. (2020). Math self-efficacy and the likelihood of pursuing a STEM-based career: a gender-based analysis. *Can. J. Sci. Math. Technol. Educ.* 20, 538–556. doi: 10.1007/s42330-020-00105-7

Heaverlo, C. A., Cooper, R., and Lannan, F. S. (2013). STEM development: predictors for 6th-12th grade girls' interest and confidence in science and math. *J. Women Minorities Sci. Eng.* 19, 121–142. doi: 10.1615/JWomenMinorScienEng.2013006464

Pereault, A., Franz-Odendaal, T. A., Langelier, E., Farenhorst, A., Mavriplis, C., and Shannon, L. (2018). "Analysis of the distribution of females and males in STEM fields in Canada," in *NSERC Chairs Women in Science and Engineering Network.* (Halifax, Canada: NSERC Network of Chairs for Women in Science and Engineering).

Steinke, J. (2017). Adolescent girls' STEM identity formation and media images of STEM professionals: considering the influence of contextual cues. *Front. Psychol.* 8:716. doi: 10.3389/fpsyg.2017.00716

Wilk, R., Ali, N., England, S. J., and Lewis, K. E. (2018). Using Zebrafish to bring hands-On laboratory experiences to urban classrooms. *Zebrafish* 15, 156–178. doi: 10.1089/zeb.2017.1503

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 22 December 2022
DOI 10.3389/fpsyg.2022.951309

OPEN ACCESS

EDITED BY
Marta Evelia Aparicio-Garcja,
Complutense University of Madrid,
Spain

REVIEWED BY
Nayssan Safavian,
University of California, Irvine,
United States
Heike Itzek-Greulich,
Thomas Morus Secondary School,
Germany

*CORRESPONDENCE
Janica Vinni-Laakso
✉ janica.vinni-laakso@helsinki.fi

SPECIALTY SECTION
This article was submitted to
Gender, Sex and Sexualities,
a section of the journal
Frontiers in Psychology

RECEIVED 23 May 2022
ACCEPTED 01 December 2022
PUBLISHED 22 December 2022

CITATION
Vinni-Laakso J, Upadyaya K and Salmela-Aro K (2022) Associations between adolescent students' multiple domain task value-cost profiles and STEM aspirations. *Front. Psychol.* 13:951309.
doi: 10.3389/fpsyg.2022.951309

Associations between adolescent students' multiple domain task value-cost profiles and STEM aspirations

Janica Vinni-Laakso*, Katja Upadyaya and Katariina Salmela-Aro

Minds Hub Research Group, Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

According to the modern expectancy-value theory, students' task values may differ across domains, manifesting as varying motivational patterns. In middle school, students' motivation becomes increasingly apparent and may direct their future occupational aspirations. Using a person-oriented approach, this study examines students' self-concept, and positive and negative task values (i.e., utility value, intrinsic value, and emotional cost) across Finnish language, math, biology, and physics, and the stability of the identified profiles. Further, the associations of the profiles with students' subsequent academic achievement and math and natural science, technology, engineering, and mathematics (STEM)/health science STEM aspirations, and gendered effects were examined. Longitudinal data was collected through Grades 7 to 9 in 21 middle schools in Helsinki, Finland ($N = 1{,}309$, $N = 1{,}179$, $N = 818$, respectively; age 13–15 years; 55.9% female). Latent profile analysis (LPA) identified four task value profiles in Grades 7 and 8: *Low motivation high cost STEM* (13%/13%) showed low task values with high cost, especially in math and physics; *High motivation low cost STEM* (7%/8%) showed the highest task values with the lowest cost, especially in math and physics; *High motivation high cost* (18%/17%) showed high task values and cost across domains; and *Moderate motivation and cost* (62%/62%) showed moderate task values and cost across domains. The latent transition analysis identified *Moderate motivation and cost* as the most stable profile across 2 years. In comparison to the other profiles, students with a *Low motivation high cost STEM* profile were less likely to have STEM aspirations in Grade 9. These results suggests that majority of middle school students are highly to moderately motivated in various domains, however, some students simultaneously experience high cost. It may reflect the increasingly difficult courses and study demands in middle school.

KEYWORDS

task values, cost, self-concept, gender, STEM aspirations, latent transition analysis

1 Introduction

Globally, there is a topical concern focusing on the increasing mismatch between the growing need for skilled labor in science, technology, engineering, and mathematics (STEM) fields and the low appeal of these areas of study and their related careers for youth (Tytler, 2014; Martin et al., 2016; OECD, 2016). In particular, attracting women and minorities to STEM-related fields has been challenging (Homer et al., 2014; National Science Foundation [NSF], 2017). Boosting STEM studies and careers among both women and men is required in order to build a more skillful workforce that is responsive to future labor market needs. In addition, researchers, educators, and policymakers should help narrow the gender gap in the STEM fields, as their actions could have multiple effects that would improve society as a whole. Ensuring that both women and men are better equipped to secure steady and well-paid jobs would ensure social mobility, advance STEM research and innovation, and reduce the risk of social exclusion for women and minorities. Many (inter)national initiatives and programs have been pursued to enhance this goal by increasing awareness of the education and career possibilities in the STEM fields and enhancing students' motivation in science (e.g., UNESCO, 2020). To understand students' educational and occupational choices, and the gendered effects, researchers have also studied the formation and development of students' science motivation and their aspirations in STEM education and careers (see, e.g., Potvin and Hasni, 2014; Wang and Degol, 2017 for reviews). In particular, the research addressing the roles of students' self-concept, interest, expectations, and achievement as the main contributors to STEM aspirations has gained a vast amount of scientific attention (e.g., for review see Watt, 2016; Guo et al., 2018; Kang et al., 2019). Utilizing a number of these constructs, we study the formation and constancy of students' task value profiles and how they predict subsequent STEM aspirations.

1.1 Expectancy-value theory

In this study, we draw on the expectancy-value framework (Eccles et al., 1983) to investigate students' task motivation in middle school. According to the expectancy-value theory, students' motivation can be divided into ability beliefs and expectancies, and subjective task values (Eccles and Wigfield, 2020). Ability beliefs and expectations relate to questions such as "Can I do this task?" (referred to here as domain specific self-concept), whereas subjective task values provide an answer to questions such as "Do I want to do this task?" Both aspects of motivation are important, as they are often associated with student achievement as well as achievement-related choices and career aspirations (for reviews see Wigfield and Cambria, 2010; Watt, 2016). However, task values are particularly important for student achievement and learning: regardless of their self-concept, a student may not engage with learning or accomplish different tasks if they do not also value the subject or activity (Ryan and Deci, 2000). Task values are further divided into intrinsic, attainment, and utility values and costs. Intrinsic value refers to students' subjective interest and the inherent enjoyment they experience when involved in a task (Eccles and Wigfield, 2020). Attainment value describes the importance of doing well in a given task, and utility value refers to the task's future relevance, or how demonstrating competence in the current task/domain will benefits one's future aspirations or career (Eccles and Wigfield, 2020).

The costs, in turn, are divided into the following categories: the demands associated with investing the significant effort required to succeed in a task (effort cost), the choices involved in setting aside other interesting/useful/important options in order to engage in a task (opportunity cost), and the psychological experiences (e.g., emotional exhaustion or stress) related to learning or completing a task (Wigfield and Eccales, 2020). To date, task value research has primarily focused on the positive values, and the perceived costs have been neglected (Flake et al., 2015), especially in longitudinal settings (Wigfield and Cambria, 2010). Positive task values and self-concept typically promote student motivation while perceived costs have been identified as a hinderance (Barron and Hulleman, 2015) in the same domain. High costs have been associated with low self-concept (e.g., Vinni-Laakso et al., 2019), interest, and academic achievement (Perez et al., 2014; Barron and Hulleman, 2015; Flake et al., 2015). High perceived costs may also lead to procrastination, avoidance behavior (Jiang et al., 2018), and impaired psychological academic wellbeing (Watt et al., 2019; Tuominen et al., 2020). Somewhat controversially, several studies have positively associated perceived cost with positive task values (Gaspard et al., 2019; Lee et al., 2022). This finding implies that high positive task values, aka high motivation, can increase the effort a student will expend in their studies, and/or they might be more willing to engage in a particular task over other valued alternatives. Emotional cost may also accompany high motivation: the stress associated with academic achievement in a given domain potentially leads to a student placing a high value on that domain. Thus, cost in the expectancy-value model could be more complex than previously assumed (Eccles et al., 1983), and it may uniquely contribute to student motivation (Barron and Hulleman, 2015).

This study aims to clarify the role of cost in student motivation by examining task value patterns in middle school. We focus on emotional cost with self-concept, intrinsic value, and utility value to gain an understanding of how negative emotional experiences interact with positive task values and contribute to motivation profiles across four domains: Finnish language, math, biology and physics.

1.2 Motivation profiles and stability

The decline of student motivation in science and mathematics has been identified in large-scale assessments, such as the Program for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMMS) (Martin et al., 2016; OECD, 2016). However, studies that methodologically examine student motivation only at the mean level cannot capture individual differences between students nor identify possible subgroups. Therefore, a number of studies have employed person-oriented approaches to research students' motivational beliefs across domains and reveal their study-related task value patterns. Prior research has found relatively similar profiles using variety of positive task value facets, namely interest, utility value and attainment value with self-concept. The profiles found often reflect high motivation with high self-concept and task values, moderate motivation with moderate self-concept and task values, and low motivation with relatively low self-concept and task values in all domains, but also mixed motivation with high self-concept and task values in some domains that are accompanied by low self-concept and/or task values in other domains (e.g., Chow and Salmela-Aro, 2011; Viljaranta et al., 2016; Guo et al., 2018; Lazarides et al., 2021; Oppermann et al., 2021). From here on, we use the terms high/moderate/low motivation or mixed motivation in profile names to refer to the relative levels of self-concept and interest/utility/attainment value in the studied samples. However, these studies have only examined positive task values and self-concept, and they have excluded cost. The few cross-sectional studies that have examined motivation patterns with cost in math have identified different profiles of students' motivation and cost; these studies have depicted high/low success expectations, utility values, and cost (Hodis and Hodis, 2020), which have reflected overall differences (e.g., low, average, high motivation) in students' task values and cost profiles (see also Fryer and Ainley, 2019). Only a few studies have identified more specific nuances in students' motivation profiles when low motivation is associated with high cost (Gaspard et al., 2019; Watt et al., 2019; Lee et al., 2022). These results have indicated that while some students with high cost will disengage from school and learning, other students may in fact orient toward high academic achievement (Conley, 2012; Watt et al., 2019; Tuominen et al., 2020). Moreover, studies that included positive and negative value beliefs (i.e., cost) also identified a moderate motivation profile that is characterized by an average level of task values and cost (Gaspard et al., 2019; Perez et al., 2019; Watt et al., 2019; Lee et al., 2022). This finding implies that costs do not function in isolation, and students may simultaneously consider that a domain is interesting and useful while engaging in study and experiencing the costs.

The person-oriented studies that have included costs have generally focused on specific domains, such as math (Watt et al., 2019; Hodis and Hodis, 2020), science (Perez et al., 2019; Watt et al., 2019), chemistry (Lee et al., 2022), or language (Fryer and Ainley, 2019). One study (Gaspard et al., 2019) that did examine task value-cost profiles across math and English as a second language identified two profiles characterized by mixed motivation (i.e., High language/Low math, Low language/High math) and two profiles with overall motivation (i.e., High motivation in language/math and Moderate motivation in language/math). The study found that perceived cost was positively associated with positive task values and self-concept in both domains, which resulted in the high motivation profile simultaneously indicating high cost. The finding showed that in contrast to the theoretical hypothesis of expectancy-value theory, cost does not merely serve as a barrier to motivation; instead, the hierarchical task values and cost together form the personal motivation patterns observed in different domains (Barron and Hulleman, 2015).

To understand students' educational choices and the factors that influence them, it is essential to first examine the formation of students' nuanced task value patterns during their middle school years, as during this period, task motivation begins to play a more important role in their studies. Prior research has shown that motivation profiles remain moderately stable over time (e.g., Fryer and Ainley, 2019; Lazarides et al., 2019; Oppermann et al., 2021), whereas some studies have found that profile memberships reveal noticeably clear changes (e.g., Lazarides et al., 2021). To the best of our knowledge, longitudinal person-oriented studies that include task values and cost remain unexplored. Additional research is required to gain an understanding of the stability of students' task value-cost patterns in middle school. Therefore, the aim of this study is to examine longitudinal profiles in students' task values and cost in Finnish language, math, biology, and physics in Grades 7 and 8 of middle school. In the expectancy-value literature, native languages and math have received extensive research attention as the stereotypical female and male domains (for a review see Wigfield and Eccales, 2020), whereas studies considering physics and biology are more recent and scarce. In order to examine STEM aspirations, three STEM-related domains were selected with Finnish language to project the findings of this study to the prior findings, and to examine stereotypically gendered motivational beliefs across domains.

1.3 Task motivation, achievement, and STEM aspirations

Expectancies and values often predict students' school achievement and direct their educational choices (Bong, 2001; Guo et al., 2015) and occupational aspirations (Chow et al., 2012; Guo et al., 2015, 2017, 2018). For example, high expectancies and/or self-concept and task values in math, physics, and chemistry predict students' entry to STEM education programs and further occupations in STEM fields (Bong, 2001; Jiang et al., 2020; Wille et al., 2020). In particular, students' math interest and utility values in middle school are associated with their choice of STEM major when enrolling in higher education (Guo et al., 2015). Similar results have also been reported for math-intensive STEM majors (Wille et al., 2020). In addition, science interest already appears to be relevant in the formation of elementary students' occupational STEM aspirations (Vinni-Laakso et al., 2019). Cost has also been shown to influence adolescent students' academic behaviors and outcomes. High cost is associated with lower academic performance in higher education (Flake et al., 2015) and contributes to increased intentions to withdraw from a STEM education/major in college (Perez et al., 2014). In middle and high school, high perceived cost was associated with students' adoption of avoidance goals, negative classroom affect, procrastination, intentions to divert from studying, and achievement in mathematics (Jiang et al., 2018; Jiang and Rosenzweig, 2021).

Rather than showing uniformly high levels of task values, the patterns of student motivation vary in terms of academic achievement and educational choices and reveal task values with intraindividual hierarchies that contribute differently to students' decisions and choices (Eccles and Wigfield, 2020). Task motivation patterns may affect students' academic achievement through the educational levels (Eccles et al., 1993; Wigfield et al., 1997; Guo et al., 2017) and further guide their educational choices and aspirations (Perez et al., 2014; Guo et al., 2015; Jiang et al., 2020). The students with high task values in math and science and low task values in other domains are more likely to pursue STEM fields than students with high task values in all domains in elementary (Oppermann et al., 2021) and secondary school (Chow et al., 2012; Guo et al., 2018; Gaspard et al., 2019). There has been less research, however, examining how perceived costs relate to students' motivational patterns and shape STEM pathways. The few studies that have investigated patterns of self-concept, positive task values, and cost have shown that these motivational constructs are associated with academic outcomes (Gaspard et al., 2019; Perez et al., 2019; Lee et al., 2022). Middle school students that were identified in mixed motivation profile as having a high math and low language motivation were more likely to aspire to a STEM major in college in comparison to other profiles that showed either high or moderate motivation across domains or high language and low math motivation (Gaspard et al., 2019). Similarly, college students' motivational patterns were associated with their academic achievement (Perez et al., 2019; Lee et al., 2022). In comparison to students with a very high motivation and low cost profile or a high motivation and moderate cost profile, students grouped in the moderate motivation profile with moderate self-concept, task values, and cost demonstrated lower achievement and completed fewer courses in the same academic year and also after 4 years. Significantly, these studies assessed opportunity cost and effort cost instead of emotional cost, which is the focus of the current paper.

There is currently a void in the literature of emotional cost and how it shapes task motivation and students' academic performance and outcomes. As opportunity and effort costs, also emotional cost has found to be negatively related to interest, utility and attainment value (Barron and Hulleman, 2015; Flake et al., 2015). However, emotional cost as a psychological factor relates more closely to emotion regulation and wellbeing (e.g., stress, exhaustion, anxiety), whereas opportunity cost and effort, where students evaluate how much time and effort they need to or are willing to put on a task/domain in order to succeed, are not emotionally draining. As shown, for some students high utility value and attainment value are accompanied with high emotional cost (Watt et al., 2019; Tuominen et al., 2020) and may have detrimental consequences in students' psychological academic wellbeing. It is important to bear in mind that emotional costs in academic setting may contribute to developing burnout symptoms which in turn may lead to lower academic achievement, lower educational aspirations, and even drop-out in later education (Salmela-Aro, 2017). In order to understand the role of emotional cost in task motivation and to identify possible vulnerable groups, we need to examine patterns of positive task values simultaneously with emotional cost. Here, we follow the theoretical framework in which intraindividual hierarchies of expectancies and task values across domains direct students' academic choices. In this study, we examine how middle school students' self-concept, interest, utility value, and emotional cost in the domains of Finnish, math, biology, and physics function together to predict students' academic achievement and occupational STEM aspirations. It is crucial to investigate students' motivational patterns in middle school in order to understand how they direct students' achievement choices in the transition to higher secondary education.

1.4 Gendered differences in science motivation and STEM aspirations

Studies have found gendered differences in students' task values and achievement across domains, and most frequently in languages, math, and science. It has been shown that, in comparison to girls, boys generally report higher self-concept and task values in math and science; however, girls have been shown to report higher self-concept and task values in verbal domains (e.g., Jacobs et al., 2002; Nagy et al., 2008; Watt et al., 2012; Gaspard et al., 2015; Guo et al., 2015). In general, girls show higher academic achievement across domains (Watt, 2016). Moreover, gender differences often occur in task motivation patterns, which show that in math and science, girls typically belong to the low motivation profile while boys often have a high motivation profile (Chow and Salmela-Aro, 2011; Guo et al., 2018; Gaspard et al., 2019; Oppermann et al., 2021). In addition, studies have shown that boys often report more STEM aspirations than girls (Eccles, 2011; Wang and Degol, 2013). Recently, researchers have begun to broaden the traditional STEM categories to include the math and natural sciences (incorporating physical science, technology, engineering, and mathematics) as well as the life sciences and medical sciences (see Dicke et al., 2019; Toh and Watt, 2022). Previous research has shown that girls aspire to the life science occupations more often than boys, whereas boys are more likely to express an interest in math and natural science occupations (e.g., Dicke et al., 2019; Oppermann et al., 2021; Toh and Watt, 2022). The recent STEM categorization of the math and natural science and health science domains offers a way to examine the nuanced gendered pathways toward STEM careers.

From this standpoint, the present study investigates the patterns and stability of students' task values and cost across multiple domains and their connection to later academic achievement and STEM aspirations. By focusing on both the positive and negative task values across domains, this study aims to clarify how task values and emotional cost are associated among individual students and how they form domain specific motivation patterns. In addition, this study examines the possible gendered differences in students' motivational patterns, academic achievement, and STEM aspirations.

2 The current study

2.1 The finnish education context

In Finland, students complete 1 year of compulsory kindergarten before they start school in the year they turn 7. Elementary education covers Grades 1–6, after which students enter middle school (Grades 7–9). All of the domains in middle school have a subject teacher, whereas the lower Grades 1–6 are taught by a homeroom teacher. Students in Finland are directed into a specific study path in Grade 9 when they are 16 years of age, which is relatively late compared to many other countries. The choices for secondary education follow students' educational aspirations by directing them into an academic track, a vocational track, or both. The selection of students for each school is based on students' preferences and their grade point average (GPA). In addition, when students enter high school, they need to select either the basic math track or the advanced math track, which differ in terms of the number of courses and the level of difficulty. This choice creates a critical filter for further STEM education, as without completing the advanced math studies in high school, students' options to apply for university STEM programs are limited. Thus, it is worthwhile to investigate students' task values in middle school as relevant antecedents for educational choices in high school.

2.2 Objectives

Research question 1: What motivational profiles can be identified in Grades 7 and 8 according to the level of students' interest and utility value, self-concepts of ability, and cost in Finnish language, math, biology, and physics?

> Hypothesis 1: We expected to find four motivation profiles: a high motivation profile characterized by high positive task values, and self-concept in all domains (e.g., Viljaranta et al., 2016; Gaspard et al., 2019; Lazarides et al., 2021; Oppermann et al., 2021); a low motivation profile with low positive task values, and self-concept across domains; a mixed motivation profile with high positive task values, and self-concept in math and physics and low positive task values, and self-concept in Finnish (Oppermann et al., 2021); and finally, a moderate motivation profile with average positive task values, and self-concept across domains (Gaspard et al., 2019; Perez et al., 2019). Based on the few prior studies that have addressed cost, we expected that for some students, high motivation may accompany high cost (Watt et al., 2019; Tuominen et al., 2020; Lee et al., 2022). As there is a lack of previous empirical studies, the research examining the role of cost in students' cross-domain motivation profiles was exploratory.

Research question 2: To what extent do students' profile memberships change from Grade 7 to 8?

> Hypothesis 2: Based on prior research, we expected the motivational profiles to be somewhat stable from Grade 7 to 8 (e.g., Lazarides et al., 2019; Oppermann et al., 2021). However, our hypotheses about the stability of motivational patterns were tentative given the lack of systematic longitudinal research simultaneously examining self-concept, positive task values, and cost in multiple domains.

Research question 3: Do students' motivational profiles differ in terms of their subsequent academic achievement?

> Hypothesis 3: We expected that a high motivation profile with high positive task values and self-concept and high or low cost would be associated with the highest academic achievement (Gaspard et al., 2019). In addition, we expected that a low motivation profile with low positive task values and self-concept would reflect the lowest academic achievement and be clearly differentiated from other profiles

(Perez et al., 2019). However, given that prior studies have rarely simultaneously researched self-concept, positive task values, and cost in multiple domains, our hypotheses regarding motivational patterns predicting achievement remained tentative.

Research question 4: To what extent do the identified motivational profiles differ in terms of students' STEM aspirations?

Hypothesis 4: We expected that a high motivation profile with high positive task values and self-concept across domains and/or high motivation in math and physics (e.g., Chow et al., 2012; Guo et al., 2018; Oppermann et al., 2021) would be associated with the highest occupational STEM aspirations. Again, our hypotheses about the joint cross-domain motivational patterns predicting STEM aspirations were empirical.

Research question 5: Do students' motivational profile memberships, academic achievement, and STEM aspirations differ in terms of gender?

Hypothesis 5: We expected that girls would be more likely to have a high motivation profile with high positive task values and self-concept across domains (e.g., Chow et al., 2012; Watt et al., 2019; Oppermann et al., 2021) while boys would be more likely to have a low motivation profile across domains (Watt et al., 2019; Oppermann et al., 2021) and/or a math-motivated profile (Chow et al., 2012; Guo et al., 2018; Oppermann et al., 2021). We also expected girls to show higher academic achievement across the measured domains (Watt, 2016) and have more health science STEM aspirations than boys, and we expected boys to report more math and natural science STEM aspirations than girls (Dicke et al., 2019; Toh and Watt, 2022).

3 Materials and methods

3.1 Participants and procedure

The data was collected from students in Grades 7–9 (N = 1,309, N = 1,179, N = 818, respectively; age 13–15; 55.9% female) in a total of 21 middle schools in the Helsinki metropolitan area during the spring semesters of the years 2014–2016. Population in Finland is homogeny regarding the racial variation where 5% of the population had a foreign background in year 2021 (Suomen Virallinen Tilasto [SVT], 2022).[1] Moreover, families' socioeconomic (SES) variation is minimal as low income families are supported by social welfares. Thus, collecting information on family's SES from students' is challenging, resulting that the data only include students' self-report information of their parent working/not working. Snowball sampling strategies were used to include new students and schools each year. Students filled in paper-based self-reports during class. Active parental consents were obtained from all participating students. The Education Division of the city of Helsinki pre-examined the research plan and gave permission to conduct the study.

3.2 Measures

3.2.1 Subjective task values

An adapted task value scale (Eccles et al., 1983) was used to assess students' subjective task values and included *Utility value* ("*The subject is useful*"), *Interest* ("*The subject is interesting*"), *Self-concept* ("*I am good at the subject*"), and *Cost* ("*The subject exhausts me*") for Finnish language, mathematics, biology, and physics on a seven-point Likert scale (1 = Not at all, 7 = Very much). Scale reliability estimates (i.e., Cronbach's alpha) cannot be provided because of the one-item measure for the subjective task values.

3.2.2 Occupational aspirations

In the third data collection wave, students' occupational aspirations were measured with an open-ended question: "What kind of work would you like to do when you grow up?" The students' responses were first coded into occupational fields based on International Standard Classification of Occupations, 2008 (ISCO-08) endorsed by the Governing Body of the International Labor Organization (ILO). These classifications were then further divided into (1) non-STEM, (2) health science occupations, and (3) math and natural science occupations including engineering and ICT following the OECD STEM classification used in OECD (2016) (see Results, Annex A1). We used these classification criteria based on the field of occupation, and did not divide students occupational aspirations by the level of education (professional and assistant). As an exception for ISCO-08 coding, a psychologist was considered as a health profession and categorized as health and medical science occupations not as a law/culture/social sciences. Students most frequent answers coded as Math and natural science STEM were an architect, an engineer, and a programmer, whereas the most frequent occupations coded as Health science STEM were a doctor, a veterinarian, psychologist, and a nurse. The most frequent answers coded as non-STEM occupation were a lawyer, a teacher, an entrepreneur, a pilot, a police officer, a dancer, and

1 tilastokeskus.fi

an actor (see Appendix for the full list of named occupational aspirations). We admit that STEM categorization were in some cases ambiguous (for example a researcher can be in the various fields but are here coded as math and natural science STEM), and sometimes students answers were difficult to interpret as in the case "something related to art." The encoding followed the coding scheme and was completed by two persons separately. The majority of the responses ($N = 413$) were coded as non-stem occupations ($n = 257$; 62.1%) while 27.5% of the responses were coded as health science STEM ($n = 114$) and only 10.4% of the responses as math and natural science STEM ($n = 43$). Based on these classifications, three dummy variables were created: (1) Math and natural science STEM vs. other fields; (2) Health science STEM vs. other fields, and (3) Combined STEM including both math and natural science and health science STEM vs. other fields (see **Table 2**).

3.2.3 Achievement data

Students' achievement data in Finnish language, math, biology, and physics were retrieved from the registry of the Finnish National Agency for Education. The achievement data were further used as a mean sum score of general GPA in the analyses because it has been shown that academic performance has high correlations across domains in basic education, meaning that students who perform well in math most often perform well in also language (Kupiainen et al., 2014).

3.2.4 Background information

The background information collected in the questionnaire included gender (0 = girl, 1 = boy) and age (i.e., date of birth).

3.3 Analytical strategy

In preliminary analysis the descriptive data and correlations of the study variables were examined (see **Table 1**). Latent profile analysis offer a way to detect different motivation patterns of self-concept and positive and negative task values that might vary across multiple domains. The strength of this analysis is to reveal subgroups in student population that would remain hidden in the average mean level scrutiny. To examine RQ1, the LPAs were conducted separately for each time point including task values across Finnish language, math, and physics. The established profile solutions were based on the akaike information criterion (AIC), the bayesian information criterion (BIC), the sample-size-adjusted Bayesian information criterion (aBIC), and the adjusted Lo-Mendell-Rubin likelihood ratio test (LMR LRT) to examine the difference in the model fit (Nylund et al., 2007). A model with lower AIC, BIC, and aBIC values was considered the best fit to the data. Classification quality was considered in terms of entropy and average class probability for the most likely class membership. In addition, the theoretical interpretation of the profiles and the number of cases in the profiles were considered in the model selection where profiles $n > 5\%$ of the sample are typically not considered as relevant subgroups (see the guidelines provided by Marsh et al., 2009) (see **Table 3** for model fit criteria in LPA).

To examine RQ2, stability and change in the students' latent profile membership were examined with latent transition analysis (LTA) (Asparouhov and Muthen, 2014). This was done by first testing measurement invariance with longitudinal constraints across the measurement points including profile similarity (Model 1–5), and second, by estimating the transition with saved class probabilities (Model 6). The advantage of using LTA in estimating the transition is that it uses the probability in estimation; thus, instead of fixed groups of students, the uncertainty of the profile membership in each time point is considered (Asparouhov and Muthen, 2014).

After the transition analysis, the auxiliary models were estimated using a manual R3STEP approach (Asparouhov and Muthen, 2014), which produced outputs that could be interpreted as multinomial logistic regression. We first tested gender moderation (Model 7 with free relations and Model 8 with equal relations) in order to later examine gendered effects reliably, and then estimated how gender predicts profile membership to examine RQ5. After this, we examined RQ3 by predicting students' academic achievement by their GPA in matching domains a year later (Model 9a and 9b).

Finally, to examine RQ4, students' STEM aspirations in Grade 9 were predicted with Grade 8 profiles (Model 10a and 10b); these analyses were also performed separately for aspirations coded as health science STEM (Model 11a and 11b) and math and natural science STEM (Model 12a and 12b).

All the models were first estimated with direct effects without gender as a covariate (Model a), and then gender was added to the models as a covariate to estimate the gendered effect in order to answer RQ5 (Model b). All models were estimated using Mplus 8.6 (Muthen and Muthen, 2018) and are presented in **Table 4**.

This project used a snowball strategy to recruit the sample; new students were included each year to compensate for the loss of previous-wave students. Of the $N = 1{,}702$ students, 768 were present in both the Grade 7 and 8 measurement points. Little's MCAR test showed that data was not missing completely at random (Chi-Square = 4,458.804 DF = 4,262, $p = 0.018$). Therefore, all models were estimated using the robust maximum likelihood estimator (MLR) with full information maximum likelihood (FIML) to handle the missing data; all the available information was used to maximize the sample size and achieve reasonable generalizability.

TABLE 1 Descriptive data and correlations of the study variables.

	Grade 8															
	Finnish				Math				Biology				Physics			
	Utility	Interest	SC	Cost	Utility	Interest	SC	Cost	Utility	Interest	SC	Cost	Utility	Interest	SC	Cost
Grade 7																
Finnish utility		0.52**	0.38**	−0.05	0.41**	0.18**	0.12**	0.10**	0.42**	0.22**	0.23**	0.03	0.34**	0.16**	0.13**	0.07*
Finnish interest	0.49**		0.54**	−0.17**	0.20**	0.29**	0.14**	0.03	0.30**	0.35**	0.27**	0.01	0.25**	0.29**	0.22**	−0.03
Finnish SC	0.42**	0.53**		−0.26**	0.23**	0.25**	0.40**	−0.05	0.22**	0.25**	0.43**	−0.05	0.20**	0.19**	0.36**	−0.04
Finnish cost	−0.19**	−0.31**	−0.35**		0.01	−0.01	−0.05	0.53**	0.05	−0.01	−0.07*	0.57**	0.07*	0.07*	−0.03	0.52**
Math utility	0.34**	0.21**	0.27**	−0.05		0.50**	0.40**	−0.12**	0.44**	0.26**	0.26**	−0.02	0.59**	0.32**	0.32**	−0.06*
Math interest	0.21**	0.34**	0.24**	−0.05	0.51**		0.71**	−0.36**	0.31**	0.44**	0.35**	−0.09**	0.43**	0.62**	0.52**	−0.21**
Math SC	0.15**	0.15**	0.36**	−0.06*	0.38**	0.66**		−0.40**	0.20**	0.27**	0.45**	−0.08**	0.33**	0.47**	0.66**	−0.23**
Math cost	−0.05	−0.05	−0.08**	0.38**	−0.16**	−0.40**	−0.45**		−0.05	−0.10**	−0.11**	0.61**	−0.13**	−0.20**	−0.28**	0.74**
Biology utility	0.39**	0.31**	0.27**	−0.06*	0.41**	0.35**	0.19**	−0.06*		0.61**	0.44**	−0.06*	0.66**	0.37**	0.26**	−0.04
Biology interest	0.28**	0.37**	0.25**	−0.08**	0.24**	0.37**	0.21**	−0.09**	0.62**		0.64**	−0.21**	0.40**	0.50**	0.35**	−0.09**
Biology SC	0.24**	0.29**	0.41**	−0.15**	0.23**	0.29**	0.36**	−0.09**	0.48**	0.66**		−0.27**	0.35**	0.37**	0.55**	−0.09**
Biology cost	−0.09**	−0.12**	−0.10**	0.48**	−0.03	−0.05	−0.03	0.45**	−0.14**	−0.30**	−0.30**		−0.03	−0.06*	−0.11**	0.66**
Physics utility	0.32**	0.25**	0.25**	0.01	0.50**	0.42**	0.31**	−0.11**	0.62**	0.42**	0.38**	−0.06		0.60**	0.50**	−0.16**
Physics interest	0.17**	0.31**	0.22**	−0.00	0.32**	0.53**	0.41**	−0.21**	0.37**	0.53**	0.38**	−0.11**	0.61**		0.70**	−0.28**
Physics SC	0.14**	0.19**	0.35**	−0.05	0.31**	0.41**	0.54**	−0.23**	0.31**	0.35**	0.52**	−0.09**	0.53**	0.68**		−0.35**
Physics cost	0.05	−0.04	0.00	0.39**	−0.08**	−0.14**	−0.12**	0.55**	−0.03	−0.10**	−0.11**	0.60**	−0.13**	−0.31**	−0.31**	
Longitudinal corr.																
Finnish utility	0.44**	0.33**	0.22**	−0.13**	0.14**	0.10**	0.00	0.04	0.19**	0.15**	0.11**	−0.06	0.12**	0.06	0.04	0.01
Finnish interest	0.27**	0.50**	0.33**	−0.17**	0.09*	0.11**	0.05	0.02	0.14**	0.22**	0.14**	−0.07	0.12**	0.15**	0.14**	−0.01
Finnish SC	0.20**	0.37**	0.49**	−0.25**	0.16**	0.19**	0.22**	−0.08*	0.12**	0.19**	0.24**	−0.14**	0.12**	0.14**	0.19**	−0.07
Finnish cost	−0.06	−0.15**	−0.22**	0.37**	−0.02	0.01	−0.06	0.15**	−0.01	0.02	−0.02	0.18**	0.09*	0.06	−0.02	0.12**
Math utility	0.20**	0.14**	0.09*	−0.04	0.39**	0.32**	0.25**	−0.15**	0.19**	0.15**	0.08*	−0.03	0.23**	0.19**	0.14**	−0.03
Math interest	0.14**	0.19**	0.16**	−0.02	0.35**	0.57**	0.52**	−0.29**	0.23**	0.26**	0.20**	−0.04	0.30**	0.38**	0.33**	−0.06
Math SC	0.08*	0.10**	0.22**	−0.05	0.31**	0.51**	0.67**	−0.39**	0.13**	0.22**	0.25**	−0.08*	0.24**	0.35**	0.39**	−0.11**
Math cost	−0.02	−0.06	−0.08*	0.17**	−0.16**	−0.32**	−0.38**	0.42**	−0.10**	−0.09*	−0.07	0.14**	−0.10**	−0.16**	−0.21**	0.21**

(Continued)

10.3389/fpsyg.2022.95309

TABLE 1 (Continued)

	Grade 8															
	Finnish				Math				Biology				Physics			
	Utility	Interest	SC	Cost	Utility	Interest	SC	Cost	Utility	Interest	SC	Cost	Utility	Interest	SC	Cost
Biology utility	0.20**	0.19**	0.10**	−0.02	0.17**	0.23**	0.13**	−0.01	0.44**	0.41**	0.30**	−0.04	0.35**	0.24**	0.21**	−0.01
Biology interest	0.14**	0.22**	0.12**	−0.01	0.16**	0.25**	0.17**	−0.04	0.41**	0.57**	0.43**	−0.18**	0.28**	0.27**	0.25**	−0.07
Biology SC	0.10**	0.17**	0.18**	−0.02	0.13**	0.22**	0.29**	−0.08*	0.31**	0.47**	0.53**	−0.19**	0.25**	0.26**	0.30**	−0.03
Biology cost	−0.05	−0.11**	−0.09*	0.23**	−0.02	−0.06	−0.07*	0.23**	−0.10**	−0.16**	−0.19**	0.35**	−0.00	−0.04	−0.06	0.23**
Physics utility	0.18**	0.15**	0.07	0.03	0.30**	0.31**	0.23**	−0.14**	0.32**	0.29**	0.21**	−0.04	0.47**	0.37**	0.31**	−0.07
Physics interest	0.14**	0.18**	0.15**	0.04	0.26**	0.39**	0.36**	−0.19**	0.27**	0.29**	0.26**	−0.05	0.43**	0.48**	0.46**	−0.13**
Physics SC	0.09*	0.15**	0.22**	−0.02	0.30**	0.40**	0.49**	−0.25**	0.20**	0.28**	0.33**	−0.08*	0.36**	0.42**	0.50**	−0.16**
Physics cost	−0.01	−0.06	−0.08*	0.19**	−0.13**	−0.20**	−0.22**	0.34**	−0.08*	−0.05	−0.08*	0.21**	−0.11**	−0.15**	−0.22**	0.30**
Grade 7																
Mean	5.46	3.93	5.09	3.53	5.82	4.20	4.83	4.19	4.78	4.18	4.74	3.74	4.69	4.01	4.37	4.13
SD	1.52	1.72	1.32	1.83	1.39	1.88	1.67	1.96	1.52	1.85	1.42	1.78	1.64	1.94	1.57	1.82
N	1,278	1,251	1,274	1,239	1,265	1,250	1,255	1,221	1,268	1,250	1,252	1,208	1,244	1,223	1,214	1,195
Grade 8																
Mean	5.52	4.24	5.19	3.65	5.73	4.39	4.74	4.32	4.81	4.29	4.82	3.87	4.75	4.05	4.37	4.41
SD	1.58	1.83	1.43	1.96	1.50	1.98	1.77	2.03	1.62	1.90	1.50	1.87	1.80	2.05	1.76	1.93
N	1,157	1,143	1,150	1,119	1,151	1,140	1,148	1,118	1,148	1,139	1,145	1,112	1,149	1,139	1,141	1,116
Range	1–7	1–7	1–7	1–7	1–7	1–7	1–7	1–7	1–7	1–7	1–7	1–7	1–7	1–7	1–7	1–7

Cross-sectional correlations under the diagonal for Grade 7 and above the diagonal for Grade 8; longitudinal correlations are under the cross-sectional estimates. SC, self-concept; GPA, grade point average of the measured subject domains; SD, standard deviation of the estimate. $^{**}p < 0.01$, $^{*}p < 0.05$.

10.3389/fpsyg.2022.951309

TABLE 2 Descriptive data of achievement and occupational aspirations.

	n	M	SD
GPA 8	1,302	8.14	1.07
GPA 9	1,219	8.19	1.14
	Frequency of named aspirations for $n = 413$ (full sample for open-answered question)		
	n	%	Gender ratio per aspiration (female in %)
Health STEM	114	27.5	23.5
Other STEM	43	10.4	2.9
STEM (combined)	155	35.5	26.4
non-STEM	257	62.1	37.8

M, mean; SD, standard deviation of the mean estimate.

4 Results

4.1 Motivation profiles

Four similar task value profiles were identified in Grade 7 and 8 (see Figure 1 for centered mean differences). *Low motivation high cost STEM* (13% t1; 13% t2) showed the lowest utility value, interest, and self-concept with the highest cost across domains, and notably low interest and high cost in math and physics. *High motivation low cost STEM* (7% t1; 8% t2) was the smallest profile in both time points and showed the highest utility value, interest, and self-concept with the lowest cost across domains, and again particularly in math and physics. *High motivation high cost* (18% t1; 17% t2) also showed high utility value, interest, and self-concept, accompanied with relatively high cost across domains. *Moderate motivation and cost* (62% t1; 62% t2) was the largest profile and showed moderate task values and cost across domains. The last two profiles showed no clear differences between domains.

4.2 Stability of the profile memberships and transition patterns

Latent transition analysis revealed that students were most likely to move to a *Moderate motivation and cost* profile or remain in their original profile from time 1 to time 2. *Moderate motivation and cost* was the largest and most stable profile across Grade 7 and 8 (transition probability 0.65). The *High motivation low cost STEM* profile was the least stable (transition probabilities 0.26), and *Low motivation high cost STEM* and *High motivation high cost* were slightly more stable profiles (transition probabilities 0.34 and 0.32, respectively) (see Table 5 for details). The transition patterns (Figure 2) indicated that the most frequent transitions across profiles were between *High motivation high cost* and *Moderate motivation and cost* (P3→P4 10.9% and P4→P3 10.5%) as well as between *High motivation low cost STEM* and *Moderate motivation and cost* (P1→P4 8.5% and P4→P1 8.9%). Students that were identified in the smallest and least stable profile *High motivation low cost STEM* were more likely to transition to the *Moderate motivation and cost* profile (3.9%). The percentages provided in the study represent the proportion of students in the total sample ($N = 1{,}702$ using FIML; Details are shown in Table 5).

4.3 Differences in academic achievement

Students' profile memberships in Grades 7 and 8 predicted their academic achievement a year later; in addition, statistically significant differences in the future achievement of the profiles were found. Academic achievement was lowest in the *Low motivation high cost STEM* profile and highest in the *High motivation low cost STEM* profile. Students' academic achievement (GPA) in Grade 8 differed between the profiles except between the two high motivation profiles: *High motivation low cost STEM* and *High motivation high cost* (Table 6). Students' achievement in Grade 9 was statistically significant between all the profiles when gender was not in the model as a covariate. However, when the gendered effect was present in the model, the differences between the profiles became non-significant and more complex: only *Low motivation high cost STEM* and *High motivation high cost* profiles remained statistically different in terms of students' academic achievement (see Table 6 for details).

4.4 Differences in STEM aspirations

Students' STEM aspirations in Grade 9 differed according to their profile membership in Grade 8. Students in the profiles *High motivation low cost STEM* and *High motivation high cost* did not differ in terms of combined STEM aspirations; in addition, the students in these two profiles were more likely to have STEM aspirations compared to students in the profiles *Low motivation high cost STEM* and *Moderate motivation and cost*. Similarly, students in the profiles *High motivation low cost STEM* and *High motivation high cost* did not differ in terms of health science STEM aspirations (coding: health science STEM vs. others), and they were more likely to have health science STEM aspirations compared to students in the profiles *Low motivation high cost STEM* and marginally significantly to *Moderate motivation and cost* when gender was added to the model. However, the significant difference between the profiles were small *High motivation low cost STEM* profile and the *Moderate motivation and cost* profile were not found in the model without gender. Only marginal profile differences were found in students' math and natural science STEM aspirations

TABLE 3 Model fit criteria of the one- to five-class solutions at T1 (Grade 7) and at T2 (Grade 8).

Model	No of profiles	#fp	LL	Scaling	AIC	BIC	aBIC	Entropy	Smallest likelihood (profile)	Size of smallest profile	LRT test
Grade 7 profile enumeration ($N = 1,309$)	1	32	−38,321.653	0.9010	76,707.306	76,872.971	76,771.322	1			
	2	49	−36,574.940	1.1353	73,247.881	73,501.555	73,345.905	0.857	0.947 (1)	40.9%	0.0000
	3	66	−36,136.697	1.2788	72,405.393	72,747.076	72,537.426	0.812	0.905 (2)	19.8%	0.0008
	4	**83**	**−35,889.798**	**1.3809**	**71,945.596**	**72,375.289**	**72,111.637**	**0.813**	**0.870 (1)**	**16.0%**	**0.1330**
	5	100	−35,621.929	1.5753	71,443.858	71,961.560	71,643.907	0.810	0.859 (1)	8.5%	0.5530
Grade 8 profile enumeration ($N = 1,176$)	1	32	−36,221.915	0.8680	72,507.831	72,670.067	72,568.423	1			
	2	49	−34,548.418	1.1735	69,194.836	69,443.259	69,287.618	0.867	0.951	38.4%	0.0000
	3	66	−33,998.197	1.2870	68,128.394	68,463.005	68,253.366	0.830	0.896	34.1%	0.0036
	4	**83**	**−33,596.625**	**1.3504**	**67,359.251**	**67,780.050**	**67,516.413**	**0.834**	**0.880 (1)**	**20.5%**	**0.0186**
	5	100	−33,200.750	1.3669	66,601.500	67,108.487	66,790.852	0.866	0.889	2.9%	0.0204

#fp, free parameters; LL, log likelihood; Scaling, log L (MLR corr. factor); aBIC, sample size adjusted BIC, LRT test, LRT test for k vs. k-1 profile. Bold values refer to the chosen profile solution.

TABLE 4 Model fit criteria for the latent transition analyses.

	#fp	LL	Scaling	AIC	BIC	ABIC
Longitudinal latent profile analysis						
Model 1. Configural similarity	166	−69,486.423	1.3658	139,304.847	140,207.814	139,680.452
Model 2. Configural with residual correlations	278	−67,820.254	3.2116	136,196.507	137,708.705	136,825.532
Model 3. Dispersion similarity (fixed variances)	214	−67,841.538	1.7089	136,111.075	137,275.141	136,595.289
Model 4. Structural similarity (fixed means)	150	−67,742.317	1.4141	135,784.634	136,600.568	136,124.036
Model 5. Distributional similarity (fixed class probabilities)	147	−67,745.898	1.4185	135,785.795	136,585.410	136,118.409
Model 6. Latent transition analysis	15	−3,523.875	0.8668	7,077.749	7,159.343	7,111.689
Predictive similarity						
Model 7. Free relations with predictor (Gender)	21	−3,354.448	0.9096	6,750.895	6,864.063	6,797.350
Model 8. Equal relations with predictor (Gender)	18	−3,355.527	0.8969	6,747.054	6,844.056	6,786.873
Explanatory similarity						
Model 9a. Relations with GPA (without covariate)	25	−7,267.335	0.8565	14,584.670	14,720.659	14,641.237
Model 9b. Relations with GPA (with covariate)	28	−6,031.242	0.9284	12,118.485	12,269.375	12,180.424
Model 10a. Relations with combined STEM (without covariate)	20	−3,799.035	0.8626	7,638.070	7,746.861	7,683.324
Model 10b. Relations with combined STEM (with covariate)	21	−3,641.385	0.8702	7,324.770	7,437.938	7,371.224
Model 11a. Relations with health science STEM (without covariate)	20	−3,770.207	0.8869	7,580.414	7,689.205	7,625.668
Model 11b. Relations with health science STEM (with covariate)	21	−3,595.942	0.8875	7,233.885	7,347.053	7,280.340
Model 12a. Relations with math and natural science STEM (without covariate)	20	−3,616.361	1.0344	7,272.721	7,381.512	7,317.975
Model 12b. Relations with math and natural science STEM (with covariate)	21	−3,447.148	1.0472	6,936.296	7,049.464	6,982.751

#fp, free parameters; LL, log likelihood; Scaling, log L (MLR corr. factor); ABIC, sample size adjusted BIC.

(coding: math and natural science vs. others): in the model without gender as a covariate, the *Low motivation high cost STEM* profile was different from the *High motivation low cost STEM* profile ($\beta = -0.137$, SE $= 0.071$, $p = 0.053$). These differences where not found in the model when gender was added as a covariate (see Table 6 for further details).

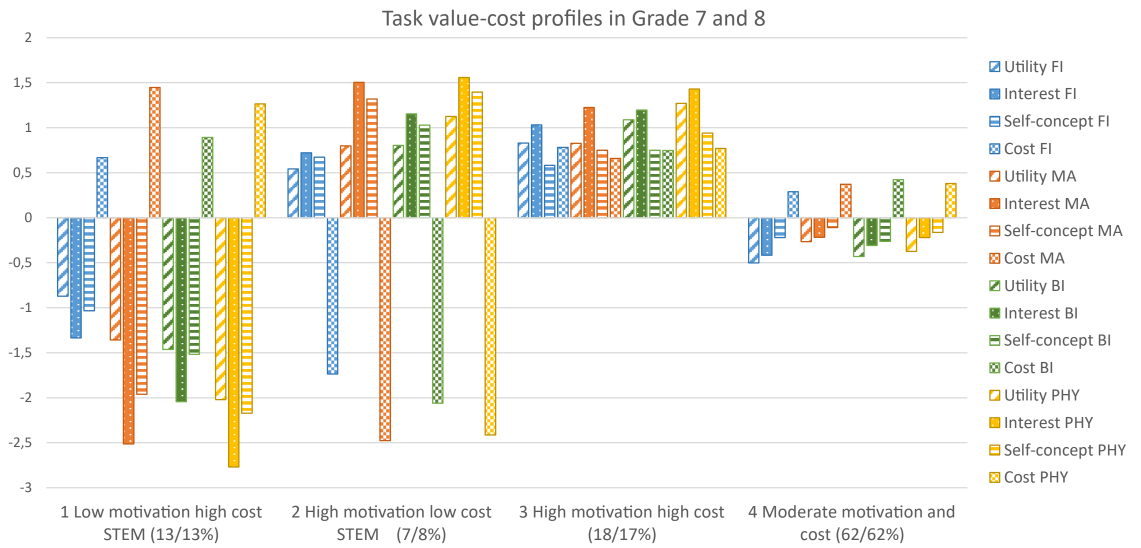

FIGURE 1
Mean levels of students' task value-cost profiles in Grade 7 and 8. Means in the figure are centered by the mean of model estimated group means; Proportion of the profiles indicate Grade 7/Grade 8 percentages; FI, Finnish language; MA, Mathematics; BI, Biology; PHY, Physics.

TABLE 5 Latent transition probabilities from grade 7 to 8.

	Transition probabilities to grade 8 profiles			
Profiles at grade 7	Low motivation high cost STEM	High motivation low cost STEM	High motivation high cost	Moderate motivation and cost
Low motivation high cost STEM	0.335	0.004	0.000	0.660
High motivation low cost STEM	0.040	0.264	0.121	0.574
High motivation high cost	0.000	0.082	0.323	0.595
Moderate motivation and cost	0.134	0.069	0.152	0.645

4.5 Gendered differences in motivational profiles and STEM aspirations

Gendered variations in the profile memberships were found in both time points. In Grade 7, more boys than girls belonged to the *Low motivation high cost STEM* profile and the boys were less likely to belong to the other profiles, namely *High motivation low cost STEM*, *Moderate motivation and cost*, and *High motivation high cost*. In Grade 8, in comparison to girls, boys also belonged to the *Low motivation high cost STEM* profile more often than the *Moderate motivation and cost* profile, while the differences between the other profiles were no longer observed (Table 7). Girls were associated with higher academic achievement in both time points compared to boys. In addition, girls had more combined STEM aspirations and were more likely to report occupational aspirations in health science STEM, while boys were more likely to report occupational aspirations in math and natural science STEM in comparison to girls (Table 8).

5 Discussion

During the middle school years, students' motivation becomes more differentiated and begins to direct their future occupational aspirations (Gaspard et al., 2017; Guo et al., 2017). Students report diverse expectancies and values: motivational patterns are formed by the intraindividual hierarchies of task values and costs that vary among students and across domains (Gaspard et al., 2019). This study contributes to the expectancy-value literature in several ways: first, by investigating the associations between the positive and negative task values simultaneously across multiple domains using a longitudinal person-oriented approach; second, by investigating the stability of the identified task value-cost profiles over time; third, by examining how the task value-cost profiles are associated with subsequent

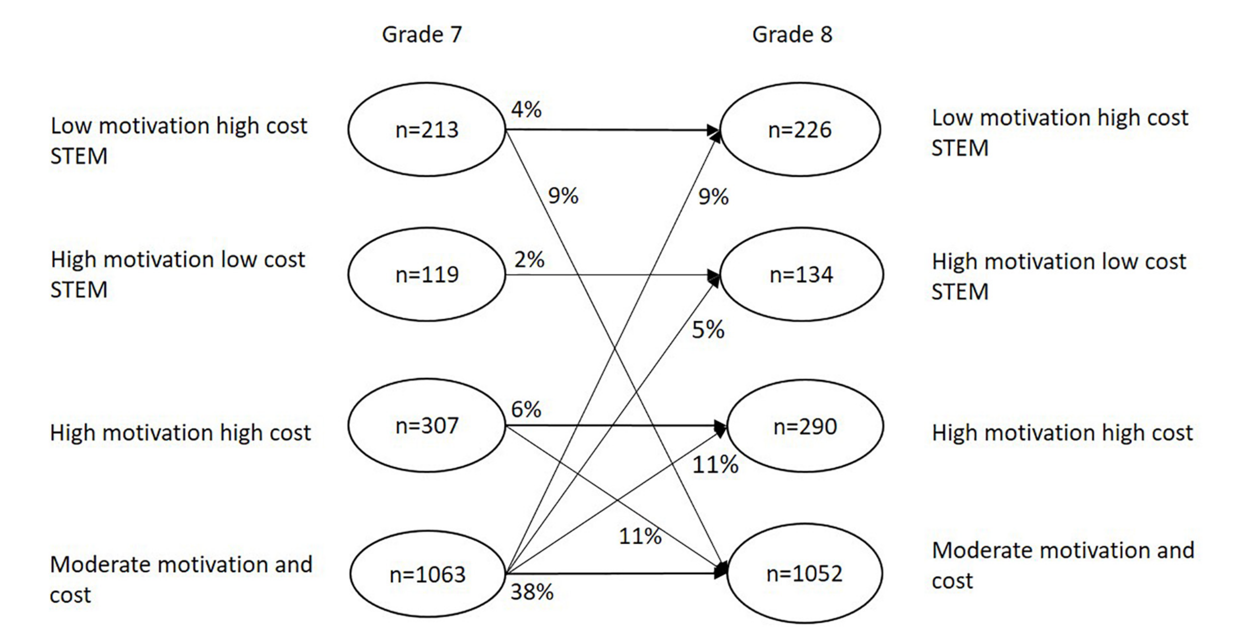

FIGURE 2
Latent transition patterns with $N = 1{,}702$ cases. Only the changes that occurred in more than 4% of the total sample ($N = 1{,}702$) with FIML estimation are depicted. All other changes are reported in **Supplementary Table 1**. The numbers in the circles refer to the final class proportions for each latent class variable based on their most likely class membership. The numbers on the arrows refer to transition probabilities for the latent class changes based on the estimated model.

TABLE 6 Task value-profiles and academic achievement and STEM aspirations.

	P1: Low motivation high cost STEM	P2: High motivation low cost STEM	P3: High motivation high cost overall	P4: Moderate motivation and cost overall
	M [SE]	M [SE]	M [SE]	M [SE]
Grade 8				
Relations with GPA (without covariate)	7.55 [0.11]	8.69^{a} [0.10]	8.54^{a} [0.08]	8.10 [0.04]
Relations with GPA (with covariate)	8.27 [0.06]	8.52^{a} [0.06]	8.47^{a} [0.05]	8.38 [0.04]
Grade 9				
Relations with GPA (without covariate)	7.62 [0.09]	8.94 [0.09]	8.50 [0.08]	8.12 [0.05]
Relations with GPA (with covariate)	8.38^{a} [0.06]	8.50^{ab} [0.06]	8.51^{b} [0.05]	8.43^{ab} [0.04]
Relations with STEM (without covariate)	0.13 [0.05]	0.55^{a} [0.09]	0.52^{a} [0.06]	0.32 [0.03]
Relations with STEM (with covariate)	0.15 [0.05]	0.59^{a} [0.09]	0.55^{a} [0.07]	0.35 [0.04]
Relations with health science STEM (without covariate)	0.14^{c} [0.05]	0.38^{ab} [0.09]	0.42^{a} [0.06]	0.24^{bc} [0.03]
Relations with health science STEM (with covariate)	0.20 [0.05]	0.50^{ab} [0.08]	0.53^{a} [0.06]	0.34^{b} [0.04]
Relations with MPECS STEM (without covariate)	0.04^{a} [0.03]	0.22^{b} [0.08]	0.10^{ab} [0.04]	0.10^{b} [0.02]
Relations with MPECS STEM (with covariate)	0.00^{a} [0.03]	0.15^{a} [0.08]	0.03^{a} [0.03]	0.05^{a} [0.02]

Means sharing the same superscript are not significantly different at $p < 0.05$. Means without the superscript accordingly significantly differ from all other profiles, marginally significant differences at $p < 0.06$ are marked with gray superscript.

academic achievement and STEM aspirations; and fourth, by examining the possible gender differences in students' task value-cost profiles, academic achievement, and STEM aspirations. In addition, this study uses more nuanced categorization to examine students' STEM aspirations in the fields of math and natural sciences, and health and medical sciences providing relevant information of gendered career aspirations.

TABLE 7 Effect of gender on latent profile membership.

	OR	SE	95% CI
Grade 7			
P1 vs. P2	0.37***	0.10	[0.22; 0.62]
P1 vs. P3	0.50**	0.11	[0.32; 0.77]
P1 vs. P4	0.47***	0.09	[0.33; 0.69]
P2 vs. P3	1.34	0.33	[0.83; 2.16]
P2 vs. P4	1.28	0.26	[0.85; 1.92]
P3 vs. P4	0.95	0.15	[0.70; 1.30]
Grade 8			
P1 vs. P2	0.60	0.16	[0.36; 1.02]
P1 vs. P3	0.65	0.15	[0.42; 1.03]
P1 vs. P4	0.54**	0.10	[0.37; 0.78]
P2 vs. P3	1.08	0.27	[0.67; 1.75]
P2 vs. P4	0.89	0.19	[0.59; 1.35]
P3 vs. P4	0.82	0.13	[0.60; 1.13]

N = 1,618. 0, girls; 1, boys; OR, odds ratios; SE, standard error; 95% CI, 95% confidence intervals. $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 8 Gendered effects on achievement and STEM aspirations.

	β	SE	*p*
Grade 8			
Gendered effect on GPA	−0.514	0.057	0.000
Grade 9			
Gendered effect on GPA	−0.444	0.063	0.000
Gendered effect on STEM	−0.097	0.048	0.043
Gendered effect on health science STEM	−0.273	0.040	0.000
Gendered effect on natural science STEM	0.155	0.036	0.000

0 = girls, 1 = boys.

5.1 Motivation profiles

Four task value-cost profiles were identified in Grades 7 and 8. *Low motivation high cost STEM* (13% t1; 13% t2) showed the lowest task values with the highest cost across all domains, but especially in math and physics. In turn, *High motivation low cost STEM* (7% t1; 8% t2) showed high task values and low cost, especially in math and physics. *High motivation high cost* (18% t1; 17% t2) showed high task values accompanied with relatively high cost across domains. *Moderate motivation and cost* (62% t1; 62% t2) showed moderate task values and cost across domains. The *High motivation low cost STEM* profile was the smallest group, whereas the *Moderate motivation and cost* was clearly the largest profile at both time points.

The results of this study supported earlier findings and confirmed our hypothesis regarding the number of profiles and the task value-cost patterns. Four profiles were identified, which is typical in person-oriented studies using task values (Chow et al., 2012; Guo et al., 2018; Lazarides et al., 2019). The task value and cost patterns also resembled the profiles that have been found in previous studies using the positive and negative aspects of the task values (Lee et al., 2022) and across math and English as the second language (Gaspard et al., 2019). The profiles *High motivation high cost* and *High motivation low cost STEM* confirmed our hypothesis that high motivation patterns would be observed with high and low cost. In addition, *Low motivation high cost STEM* exhibited the expected low motivation pattern. The profiles *High motivation low cost STEM* and *Low motivation high cost STEM* showed patterns of mixed motivation across domains and confirmed our hypothesis (Gaspard et al., 2019; Oppermann et al., 2021). Finally, the *Moderate motivation and cost* profile demonstrated the expected pattern with average task values.

In this study, over half of the students belonged to the *Moderate motivation and cost* profile, which confirms the findings of earlier studies that did not identify clearly differentiated task values and costs among groups of students (Perez et al., 2019; Watt et al., 2019). This finding indicates that the majority of middle school students are somewhat motivated to study, and they have not yet have developed highly distinguished task values in Finnish language, math, biology, and physics; in addition, middle school students feel moderately exhausted by their studies in all domains. This could be considered as a typical student in Middle school. The *High motivation profile with high cost* depict a typical high achieving student, most likely girl, who is highly motivated toward school and is determined to perform well in all domains. This profile could be in risk of studyholism and study burnout. However, two smaller groups of students report high or low positive task values especially in STEM domains depicting two opposite motivation patterns. It seems that STEM domains divide student motivation clearly into two groups where students are either highly motivated in math and physics with no perceived cost or considerably unmotivated in math and physics with high cost.

5.2 Stability of the profiles and transitions in profile membership

Latent transition analysis further revealed that *Moderate motivation and cost* was the most stable profile over time; the other profiles showed rather low stability. Previous research that has used LTA to examine patterns of students' expectancies and values has found moderately stable motivation profiles (e.g., Oppermann et al., 2021), but low stability has also been observed to some extent (Lazarides et al., 2021). However, these studies have only included the positive task values across domains. This study investigated task values and cost simultaneously across several domains, and thus provides new insights by showing that as the variation in the motivation profiles increases it may result in reduced stability over time. Moreover, middle school students undergo major developmental changes (e.g., puberty, adjustment to the school transition from primary to middle school, changes in peer relations), which may affect their academic motivation. Therefore, task motivation might be more prone to changes in middle school when internal and external frames of reference influence the hierarchies of students' expectancies and values in many subjects (see Marsh, 1990). Especially math physics become increasingly difficult in middle school resulting changes in students' self-concept, interest and utility values, and emotional cost in these domains when students proceed from grade 7 to 8. This might also explain the low stability in *High motivation low cost STEM* profile. Peer interactions affect students' self-perception and motivation, and social desirability might influence especially girls' motivation in math and physics. Additional longitudinal research is required to explore the cross-domain patterns of task values and cost.

5.3 Motivation profiles and academic achievement

Students' profile membership in Grades 7 and 8 predicted their academic achievement a year later, and the profiles differed according to students' academic achievement. As expected, the high motivation profiles, namely *High motivation low cost STEM* and *High motivation high cost*, were associated with the highest academic achievement, while the *Low motivation high cost STEM* profile was shown to have the lowest academic achievement. *Moderate motivation* profile showed moderate achievement; significantly lower than the two high motivation profiles but higher than the *Low motivation high cost STEM* profile. In Grade 8, no differences in students' GPA were found between the two high motivation profiles; however, differences were present in Grade 9. Moreover, when gender was included in the model, the differences between the profiles became non-significant and more complex: students in the *Low motivation high cost STEM* profile had a lower GPA compared to students in the *High motivation high cost* profile when students' gender was taken into account. These findings indicate that the association between student motivation and subsequent academic performance become stronger when students continue to pursue their educational path, and gender may play a role in this relationship by showing more differentiated motivation patterns and less clear achievement gaps between male and female students.

5.4 Motivation profiles and STEM aspirations

The results showed that students who reported *High motivation low cost STEM* or *High motivation high cost* and, to some extent, students with a *Moderate motivation and cost* profile had more combined STEM aspirations than students belonging to the *Low motivation high cost STEM* profile. This finding partially confirms our hypothesis that high motivation profiles in math and science and/or high motivation across domains is associated with more STEM aspirations compared to other profiles, an observation that is also in line with existing literature (Chow et al., 2012; Guo et al., 2018; Oppermann et al., 2021). Seems plausible that students in *Low motivation high cost STEM* profile who have low self-concept and hold low interest and utility value in math and physics and simultaneously experience high emotional cost in these domains result having no future career aspirations in STEM. The two high motivation profiles identified in this study did not show any differences in terms of students' STEM aspirations. Overall, only half of the students provided an answer when asked about a future occupation that they would want to pursue, and the majority of the occupations were coded as non-STEM. Health science STEM occupations were more frequently identified than careers in the math and natural science STEM fields. The low number of STEM aspirations might be the result of the non-significant findings between the profiles; it is possible that the students who indicated high motivation had already clearly established their future outlooks and thus were aware of more STEM occupations than the students who showed low overall motivation toward school.

5.5 Gendered motivation and STEM aspirations

This study showed significant gendered variation in the profile memberships at both time points. In Grade 7, male students were more likely to have a *Low motivation high cost STEM* profile and were less likely to belong to the other profiles, namely *High motivation low cost STEM*, *High motivation high cost*, and *Moderate motivation and cost*. In Grade 8, it was also more likely for a male student to report *Low motivation high cost STEM* than *Moderate motivation and cost*. The overrepresentation of boys in the low motivation profile is in line with frequently reported gender differences, as is the overrepresentation of girls in the high motivation profile (Chow et al., 2012; Oppermann et al., 2021). However, in the literature discussing expectancies and values, the majority of studies have reported higher motivation among boys in math and science (Watt, 2016), and this observation was not clearly replicated in this study. Most of the students who named a future occupation were girls. Moreover, the female students named non-STEM occupations more frequently than STEM occupations, and the majority of the STEM occupations were in health science STEM fields. The boys in this study named more math and natural science STEM occupations than the girls. These gendered STEM aspirations are also in line with the findings described in the existing literature (Dicke et al., 2019; Toh and Watt, 2022).

5.6 Practical implications and interventions

While there is significant awareness of the need to improve girls' engagement (UNESCO, NSF) in STEM fields, gender biases and stereotypes are still prevalent, creating obstacles to the recruitment and progression of girls in STEM education and careers. Results from intervention studies (e.g., Rosenzweig et al., 2020) have suggested that cost reduction and utility value interventions are both useful tools for improving students' STEM course performance. However, girls' academic achievement in middle school does not appear to be related to the limited number of female students pursuing a future in STEM education and careers; instead, a lack of interest in STEM fields and a stronger focus, in particular, on the internal hierarchies of other occupations may explain why girls rarely aspire to physical science occupations. By providing girls more knowledge and hands-on interactive STEM activities, it would be possible to promote girls' STEM motivation and aspirations (Franz-Odendaal and Marchand, in press) and positive emotions in science class (Itzek-Greulich and Vollmer, 2017). For example, intervention programs which would involve students discussing with role models (e.g., women working in STEM fields) may provide girls better insights into STEM careers and inspire girls to be more engaged in STEM domains (Franz-Odendaal and Marchand, in press). Moreover, previous studies have shown that female students often feel that they do not belong to STEM fields, leading them to pursue other than STEM careers (Aelenei et al., 2020). Interventions targeting sense of belonging and providing students collaborative tasks where they can work together for a common goal may support female students' interest in STEM fields (Aelenei et al., 2020). Motivation-emotion relationship should be better acknowledged in science education; by modifying teaching methods it may be possible to evoke positive achievement emotions and boost students' situational motivation in the science learning context (Itzek-Greulich and Vollmer, 2017). The findings of the current study do not show that girls experience more cost in math and physics, rather some girls may experience a cost associated with high motivation across domains. It is important to harness this high motivation and direct it into STEM-related fields; thus, there is a need to design interventions that would compensate for female students' missed opportunities to engage in science activities (Murphy and Whitelegg, 2006).

6 Conclusion

This study identified four profiles among students in middle school: two STEM-oriented profiles, one with high motivation and low cost and the other with low motivation and high cost, especially in math and physics, and two profiles depicting high motivation and cost across domains and moderate motivation and cost across domains. The moderate motivation profile was the largest and most stable profile across both Grades 7 and 8. Gendered variations in the profile memberships and STEM aspirations were also observed: girls were more likely to belong to the high motivation profiles or a moderate motivation profile, while more boys reported having a low motivation and high cost profile. Moreover, girls showed higher academic achievement in comparison to boys and had more life science STEM aspirations; in contrast, boys reported more STEM aspirations in the physical sciences. The results suggest that the majority of middle school students are moderately to highly motivated in various domains; however, some students simultaneously experience a high cost, which may reflect the increase in course difficulty and study-related demands in middle school.

6.1 Limitations and further research

Our longitudinal study was conducted with middle school students in Helsinki, Finland and included a relatively large number of participants. However, it should be noted that the participation of the same students varied across the time points. Most of the students recruited in Grade 7 remained in the study in Grade 8; however, in Grade 9, the data collection attrition increased and resulted in limited data on STEM aspirations. Students' future occupational aspirations were measured with an open-ended question that only yielded 413 answers that were further coded as non-STEM/STEM. The data for this study was collected in 21 middle schools from across the Helsinki metropolitan area and included students from various family backgrounds. However, as population in Finland is rather homogeny regarding race/ethnicity and socioeconomic background, a proper information of the SES was not collected. Further research is required to confirm the validity of the observations and the generalizability of the findings; for example, it would be desirable to extend the focus by including students from different Finnish cities or regions and even other countries. The use of a one-item task value measure in the data collection meant that we could not test the reliability of the scale, and this may weaken the validity of the study. However, we employed LPA to reduce the measurement error. While LPA is a useful means of identifying possible subgroups in the population, there are possible shortcomings related to the person-oriented methodology. We should bear in mind that the results of students' high/average/low level of task values are always relative to the used sample and cannot be interpret as objective information of student motivation in general. Moreover, these results might be different if the same analyses were conducted using another sample or in other population. In person-oriented techniques, such as LPA, the researcher is responsible for selecting and interpreting the final profile solution. While identifying profiles in the data can appear relatively straightforward, it can be difficult to classify a student in only one profile. In this study, we carefully followed standard guidelines (Asparouhov and Muthen, 2014) when conducting the LTA and confirmed that the results were aligned with the underlying theoretical framework and previous research.

The interaction of individual and contextual factors could be considered in future research. Collecting data on students' everyday experiences during classes may reveal the immediate interplay between interest and costs which could help researchers to understand the formation of students' more permanent motivation beliefs toward different domains and future career aspirations. It would also be beneficial to investigate students' levels of interest and their simultaneous perceptions of cost when engaged in different tasks within a domain (for example, math or science), and how the in-the-moment interplay is related to students' STEM aspirations. For educators, it would be important to understand the possibilities to influence task motivation in the classroom and inspire students to STEM. Moreover, it would be interesting to consider if friends share similar patterns of interests and costs, and even STEM aspirations. Examining joint motivation patterns within friend groups might reveal synchronous changes in students' task-values which further contribute to the formation of STEM aspirations as students proceed through the middle school years.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The longitudinal dataset contains pseudonymized identifiers of the under aged study participants. At this point of the research project the data cannot be published. Requests to access these datasets should be directed to JV-L, janica.vinni-laakso@helsinki.fi.

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 to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

JV-L performed the analytic calculations and wrote the manuscript with the help of KU and KS-A. KS-A was responsible for developing the original idea. JV-L and KU planned the modeling technique and the use of previously collected data. KS-A helped to supervise the project. All the authors discussed the results and contributed to the final manuscript.

Funding

This study was supported by the Academy of Finland grant nos. 308351 and 336138 to KS-A, the Strategic Research Council grant no. 345264 to KS-A, and to Kimmo Alho grant no. 312529.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.951309/full#supplementary-material

References

Aelenei, C., Martinot, D., Sicard, A., and Darnon, C. (2020). When an academic culture based on self-enhancement values undermines female students' sense of belonging, self-efficacy, and academic choices. *J. Soc. Psychol.* 160, 373–389. doi: 10.1080/00224545.2019.1675576

Asparouhov, T., and Muthen, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Struct. Equ. Modeling* 21, 329–341. doi: 10.1080/10705511.2014.915181

Barron, K. E., and Hulleman, C. S. (2015). *Expectancy-value-cost model of motivation*, 2nd Edn. Amsterdam: Elsevier, 503–509.

Bong, M. (2001). Between-and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals. *J. Educ. Psychol.* 93:23. doi: 10.1037/0022-0663.93.1.23

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.

Chow, A., Eccles, J., and Salmela-Aro, K. (2012). Task value profiles across subjects and aspirations to physical and IT-related sciences in the United States and Finland. *Dev. Psychol.* 48, 1612–1628. doi: 10.1037/a0030194

Conley, A. M. (2012). Patterns of motivation beliefs. *J. Educ. Psychol.* 104, 32–47.

Dicke, A. L., Safavian, N., and Eccles, J. S. (2019). Traditional gender role beliefs and career attainment in STEM: A gendered story? *Front. Psychol.* 10:1053. doi: 10.3389/fpsyg.2019.01053

Eccles, J. (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. (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

Eccles, J., 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*, ed. J. T. Spence (San Francisco, CA: W. H. Freeman), 75–146.

Eccles, J., 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: 10.1111/j.1467-8624.1993.tb02946.x

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., and Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemp. Educ. Psychol.* 41, 232–244. doi: 10.1016/j.cedpsych.2015.03.002

Franz-Odendaal, T., and Marchand, S. (in press). Girls get wise–a programming model for engaging girls+ in STEM. *Front. Psychol.* doi: 10.3389/fpsyg.2022.924943

Fryer, L. K., and Ainley, M. (2019). Supporting interest in a study domain: A longitudinal test of the interplay between interest, utility-value, and competence beliefs. *Learn. Instr.* 60, 252–262. doi: 10.1016/j.learninstruc.2017.11.002

Gaspard, H., Dicke, A., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., et al. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *J. Educ. Psychol.* 107, 663–677. doi: 10.1037/edu0000003

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., and Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemp. Educ. Psychol.* 48, 67–84. doi: 10.1016/j.cedpsych.2016.09.003

Gaspard, H., Wille, E., Wormington, S. V., and Hulleman, C. S. (2019). How are upper secondary school students' expectancy-value profiles associated with achievement and university STEM major? A cross-domain comparison. *Contemp. Educ. Psychol.* 58, 149–162.

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J., and Dicke, T. (2017). Extending expectancy-value theory predictions of achievement and aspirations in science: Dimensional comparison processes and expectancy-by-value interactions. *Learn. Instr.* 49, 81–91. doi: 10.1016/j.learninstruc.2016.12.007

Guo, J., Parker, P. D., Marsh, H. W., and Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Dev. Psychol.* 51, 1163–1176. doi: 10.1037/a0039440

Guo, J., Wang, M.-T., Ketonen, E. E., Eccles, J. S., and Salmela-Aro, K. (2018). Joint trajectories of task value in multiple subject domains: From both variable- and pattern-centered perspectives. *Contemp. Educ. Psychol.* 55, 139–154. doi: 10.1016/j.cedpsych.2018.10.004

Hodis, F. A., and Hodis, G. M. (2020). Patterns of motivation and communication in learning environments: A latent profile analysis. *Soc. Psychol. Educ.* 23, 1659–1685. doi: 10.1007/s11218-020-09599-3

Homer, M., Ryder, J., and Banner, I. (2014). Measuring determinants of post-compulsory participation in science: A comparative study using national data. *Br. Educ. Res. J.* 40, 610–636. doi: 10.1002/berj.3106

Itzek-Greulich, H., and Vollmer, C. (2017). Emotional and motivational outcomes of lab work in the secondary intermediate track: The contribution of a science center outreach lab. *J. Res. Sci. Teach.* 54, 3–28.

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

Jiang, Y., and Rosenzweig, E. Q. (2021). Using cost to improve predictions of adolescent students' future choice intentions, avoidance intentions, and course grades in mathematics and English. *Learn. Individ. Dif.* 86:101978. doi: 10.1016/j.lindif.2021.101978

Jiang, S., Simpkins, S. D., and Eccles, J. S. (2020). Individuals' math and science motivation and their subsequent STEM choices and achievement in high school and college: A longitudinal study of gender and college generation status differences. *Dev. Psychol.* 56:2137. doi: 10.1037/dev0001110

Jiang, Y., Rosenzweig, E. Q., and Gaspard, H. (2018). An expectancy-value-cost approach in predicting adolescent students' academic motivation and achievement. *Contemp. Educ. Psychol.* 54, 139–152. doi: 10.1016/j.cedpsych.2018.06.005

Kang, J., Hense, J., Scheersoi, A., and Keinonen, T. (2019). Gender study on the relationships between science interest and future career perspectives. *Int. J. Sci. Educ.* 41, 80–101. doi: 10.1080/09500693.2018.1534021

Kupiainen, S., Vainikainen, M.-P., Marjanen, J., and Hautamäki, J. (2014). The role of time on task in computer-based low-stakes assessment of cross-curricular skills. *J. Educ. Psychol.* 106, 627–638. doi: 10.1037/a0035507

Lazarides, R., Dicke, A.-L., Rubach, C., Oppermann, E., and Eccles, J. S. (2021). Motivational profiles across domains and academic choices within Eccles et al.'s situated expectancy–value theoretical framework. *Am. Psychol. Assoc.* 57, 1893–1909. doi: 10.1037/dev0001250

Lazarides, R., Dietrich, J., and Taskinen, P. H. (2019). Stability and change in students' motivational profiles in mathematics classrooms: The role of perceived teaching. *Teach. Teach. Educ.* 79, 164–175. doi: 10.1016/j.tate.2018.12.016

Lee, S. Y., Friedman, S., Christiaans, E., and Robinson, K. A. (2022). Valuable but costly? University students' expectancy-value-cost profiles in introductory chemistry courses. *Contemp. Educ. Psychol.* 69:102056. doi: 10.1016/j.cedpsych.2022.102056

Marsh, H. W. (1990). The influences of internal and external frames of reference on the formation of English and Math self-concepts. *J. Educ. Psychol.* 82, 107–116.

Marsh, H. W., Lüdtke, O., Trautwein, U., and Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Struct. Equ. Model.* 16, 191–225.

Martin, M. O., Mullis, I. V. S., Foy, P., and Hooper, M. (2016). *TIMSS 2015 international results in science*. Boston, MA: TIMSS & PIRLS International Study Center, Boston College.

Murphy, P., and Whitelegg, E. (2006). Girls and physics: Continuing barriers to 'belonging.'. *J. Curric.* 17, 281–305. doi: 10.1080/09585170600909753

Muthen, L. K., and Muthen, B. O. (2018). *Mplus user's guide*, 8th Edn. Los Angeles, CA: Muthén & Muthén. doi: 10.1016/j.lindif.2017.11.017

Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., Eccles, J. S., et al. (2008). "Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value," 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), 115–143. doi: 10.1037/11706-004

National Science Foundation [NSF] (2017). *Women, minorities, and persons with disabilities in science and engineering: 2017*. Arlington, TX: National Center for Science and Engineering Statistics.

Nylund, K. L., Asparouhov, T., and Muthen, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Struct. Equ. Model.* 14, 535–569. doi: 10.1080/10705511.2014.882690

OECD (2016). *PISA 2015 results (volume I): Excellence and equity in education, PISA*. Paris: OECD Publishing.

Oppermann, E., Vinni-Laakso, J., Juuti, K., Loukomies, A., and Salmela-Aro, K. (2021). Elementary school students' motivational profiles across Finnish language, mathematics and science: Longitudinal trajectories, gender differences and STEM aspirations. *Contemp. Educ. Psychol.* 64:101927. doi: 10.1016/j.cedpsych.2020.101927

Perez, T., Cromley, J. G., and Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *J. Educ. Psychol.* 106, 315–329. doi: 10.1037/a0034027

Perez, T., Wormington, S. V., Barger, Schwartz-Bloom, R. D., Lee, Y. K., and Linnenbrink-Garcia, L. (2019). Science expectancy, value, and cost profiles and their proximal and distal relations to undergraduate science, technology, engineering, and math persistence. *Sci. Educ.* 103, 264–286. doi: 10.1002/sce.21490

Potvin, P., and Hasni, A. (2014). Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Stud. Sci. Educ.* 50, 85–129. doi: 10.1080/03057267.2014.881626

Rosenzweig, E. Q., Wigfield, A., and Hulleman, C. S. (2020). More useful or not so bad? Examining the effects of utility value and cost reduction interventions in college physics. *J. Educ. Psychol.* 112, 166–182. doi: 10.1037/edu0000370

Ryan, R. M., and Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemp. Educ. Psychol.* 25, 54–67. doi: 10.1006/ceps.1999.1020

Salmela-Aro, K. (2017). Dark and bright sides of thriving–school burnout and engagement in the Finnish context. *Eur. J. Dev. Psychol.* 14, 337–349. doi: 10.1080/17405629.2016.1207517

Suomen Virallinen Tilasto [SVT], (2022). *Väestörakenne [verkkojulkaisu]*. Helsinki: Tilastokeskus.

Toh, L., and Watt, H. M. G. (2022). How do adolescent mathematical self-concept and values explain attainment of different kinds of STEM degrees in adulthood? *Contemp. Educ. Psychol.* 69:102057. doi: 10.1016/j.cedpsych.2022.102057

Tuominen, H., Juntunen, H., and Niemivirta, M. (2020). Striving for success but at what cost? Subject-specific achievement goal orientation profiles, perceived cost, and academic well-being. *Front. Psychol.* 11:557445. doi: 10.3389/fpsyg.2020.557445

Tytler, R. (2014). "Attitudes, identity, and aspirations toward science," in *Handbook of research on science education*, eds N. G. Lederman and S. K. Abell (New York, NY: Routledge), 82–103. doi: 10.1097/ACM.0000000000002006

UNESCO (2020). *Boosting gender equality in science and technology. A challenge for TVET programmes and careers*. Paris: United Nations Educational, Scientific and Cultural Organization

Viljaranta, J., Aunola, K., and Hirvonen, R. (2016). Motivation and academic performance among first-graders: A person-oriented approach. *Learn. Individ. Differ.* 49, 366–372.

Vinni-Laakso, J., Guo, J., Juuti, K., Loukomies, A., Lavonen, J., and Salmela-Aro, K. (2019). The relations of science task values, self-concept of ability, and STEM aspirations among Finnish students from first to second grade. *Front. Psychol.* 10:1449. doi: 10.3389/fpsyg.2019.01449

Wang, M. T., and Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy–Value perspective to understand individual and gender differences in STEM fields. *Dev. Rev.* 33, 304–340. doi: 10.1016/j.dr.2013.08.001

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

Watt, H. (2016). "Gender and motivation," in *Handbook of motivation at school*, 2nd Edn. eds K. R. Wentzel and D. B. Miele (New York, NY: Routledge), 320–339.

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., 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

Wigfield, A., and Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Dev. Rev.* 30, 1–35. doi: 10.1016/j.dr.2009.12.001

Wigfield, A., and Eccales, J. S. (2020). 35 years of research on students' subjective task values and motivation: A look back and a look forward. *Adv. Motiv. Sci.* 7, 161–198. doi: 10.1016/bs.adms.2019.05.002

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J., Freedman-Doan, C., et al. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *J. Educ. Psychol.* 89:451. doi: 10.1037/0022-0663.89.3.451

Wille, E., Stoll, G., Gfrörer, T., Cambria, J., Nagengast, B., and Trautwein, U. (2020). It takes two: Expectancy-value constructs and vocational interests jointly predict STEM major choices. *Contemp. Educ. Psychol.* 61:101858. doi: 10.1016/j.cedpsych.2020.101858

frontiers | Frontiers in Psychology

TYPE Systematic Review
PUBLISHED 05 January 2023
DOI 10.3389/fpsyg.2022.956300

Check for updates

OPEN ACCESS

EDITED BY
Davinia M. Resurrección,
Universidad Loyola Andalucía, Spain

REVIEWED BY
Jacinto Jardim,
Universidade Aberta, Portugal
Eva Cifre,
Universitat Jaume I, Spain

*CORRESPONDENCE
Sergi Fàbregues
✉ sfabreguesf@uoc.edu
Milagros Sáinz
✉ msainzi@uoc.edu

SPECIALTY SECTION
This article was submitted to Gender, Sex and Sexualities, a section of the journal Frontiers in Psychology

RECEIVED 30 May 2022
ACCEPTED 13 December 2022
PUBLISHED 05 January 2023

CITATION
Fàbregues S, Sáinz M, Romano MJ, Escalante-Barrios EL, Younas A and López-Pérez B-S (2023) Use of mixed methods research in intervention studies to increase young people's interest in STEM: A systematic methodological review. *Front. Psychol.* 13:956300. doi: 10.3389/fpsyg.2022.956300

Use of mixed methods research in intervention studies to increase young people's interest in STEM: A systematic methodological review

Sergi Fàbregues[1]*, Milagros Sáinz[2]*, María José Romano[2], Elsa Lucia Escalante-Barrios[3], Ahtisham Younas[4] and Beatriz-Soledad López-Pérez[2]

[1]Department of Psychology and Education, Universitat Oberta de Catalunya, Barcelona, Spain, [2]Internet Interdisciplinary Institute, Universitat Oberta de Catalunya, Barcelona, Spain, [3]Department of Education, Universidad del Norte, Barranquilla, Colombia, [4]Faculty of Nursing, Memorial University of Newfoundland, St. John's, IL, Canada

Introduction: Mixed methods research intervention studies integrate quantitative evaluation approaches, such as randomized controlled trials and quasi-experimental designs, with qualitative research to evaluate the effectiveness, efficacy, or other results of an intervention or program. These types of studies, which have attracted growing attention in recent years, enhance the scope and rigor of the evaluation. While various frameworks that summarize the justifications for carrying out these types of studies and provide implementation guidance have been published in the last few years in the health sciences, we do not know whether such frameworks have been properly implemented in the social and educational sciences. This review examined the methodological features and reporting practices of mixed methods intervention studies aimed at increasing young people's interest in STEM.

Methods: A systematic search was carried out in APA PsycNET, ERIC, ProQuest, Scopus, and Web of Science, and a hand search in 20 journals. We included peer-reviewed English-language articles that reported intervention studies with a quantitative component measuring outcomes specific to increasing secondary school students' interest in STEM fields, a qualitative component conducted before, during, or after the quantitative component, and evidence of integration of both components. Qualitative content analysis and ideal-type analysis were used to synthesize the findings.

Results: We found 34 studies; the majority published in the last ten years. Several patterns of mixed methods application were described in these studies, illustrating the unique insights that can be gained by employing this methodology. The reporting quality of the included studies was generally adequate, especially regarding the justification for using a mixed methods intervention design and the integration of the quantitative and qualitative components. Nonetheless, a few reporting issues were observed, such as a lack of detail in the presentation of the mixed methods design, an inadequate description of the qualitative sampling and analysis techniques, and the absence of joint displays for representing integration.

Discussion: Authors must pay attention to these issues to ensure that the insights obtained by the use of mixed methods research are effectively communicated.

KEYWORDS

mixed methods research, qualitative research, STEM, intervention, methodological review

1. Introduction

Mixed methods research (MMR) integrates quantitative and qualitative methods in a single study or sustained program of inquiry to generate a more complete understanding than is achievable with a single method (Fetters, 2020). The use of MMR has significantly increased in recent years and a variety of designs for its implementation have been proposed, each with its own aim, assumptions, procedures, and integration strategies (Creswell and Plano Clark, 2018). One of these is the MMR intervention design, which combines a quantitative evaluation design (i.e., randomized controlled trial [RCT], quasi-experimental design, non-experimental design) with qualitative research used to determine the effectiveness, efficacy, or other outcomes of an intervention or program. MMR intervention designs have received increasing attention in recent years. A growing number of methodological publications (Sandelowski, 1996; Lewin et al., 2009; O'Cathain et al., 2013; Zhang, 2014; Boeije et al., 2015; Grissmer, 2016; Johnson and Schoonenboom, 2016; Maher and Neale, 2019; Richards et al., 2019; Bouchard and Tulloch, 2020; Fetters and Molina-Azorin, 2020; Aschbrenner et al., 2022), including an entire textbook (O'Cathain, 2018), have described the ways in which designs of this type, when properly implemented, enhance the comprehensiveness, rigor, and efficiency of the intervention study.

One distinguishing feature of MMR intervention designs is their ability to transcend the limitations of RCTs in producing findings that are easily transferable to practice. Johnson and Schoonenboom (2016) summarized several of these limitations, including the inability to generalize the findings to other settings and populations and the fact that they are "performed in ideal circumstances" (p. 587), which might produce findings that might not be representative of the context of the intervention. Most of these limitations can be addressed by including qualitative research in the intervention study since this approach can help researchers to better understand the context and conditions surrounding the intervention, the contextual elements and causal mechanisms that generate the effects, how these mechanisms operate, and the differences between participants in the effects observed. By integrating qualitative research with a quantitative evaluation design, researchers can gather contextual and individual-specific knowledge about why, how, and under what conditions an intervention does or does not work. This more detailed understanding of the effects of the intervention will be critical in producing context-sensitive recommendations that can be implemented effectively in policy and practice. The qualitative phase, for example, might be used in implementation studies to assess the feasibility of an intervention and its implementation strategies, as well as to complete process and outcome evaluations (Cheng and Metcalfe, 2018; Landes et al., 2019).

Scholars working primarily in the health sciences have developed two main frameworks that describe reasons for using qualitative research in intervention studies. The first framework, the *temporal framework*, categorizes these reasons based on whether the qualitative component was implemented before, during, or following the intervention (Lewin et al., 2009; Johnson and Schoonenboom, 2016; Creswell and Plano Clark, 2018). For example, qualitative research undertaken before an intervention can aid researchers in evaluating the need for the intervention, generating hypotheses for testing in the quantitative part, and developing adequate outcome measures. The use of qualitative research during the intervention can aid researchers in determining the fidelity of the implementation methods, examining the perspectives of researchers carrying out the intervention, and identifying potential barriers and facilitators encountered by participants. After the intervention, researchers may use qualitative research to explain unexpected or non-significant quantitative findings, examine how the context may have influenced the findings, and identify research questions for further research. More recently, Maher and Neale (2019) proposed a variant of the temporal framework, called *temporal parallel purpose framework*, in which, maintaining the sequential logic of the previous frameworks, the authors classified the reasons according to whether they were related specifically to the intervention or the RCT. A second framework for using MMR intervention designs is the *Aspects of a Trial Framework*, which was developed from a review of 296 peer-reviewed health sciences articles published between 2008 and 2010 reporting qualitative research conducted with trials (O'Cathain et al., 2013). In that review, the authors were unable to use the temporal framework to code the reasons for doing qualitative research in the included studies because most of them did not provide the precise time period for the qualitative data collection. As a result, O'Cathain (2018), the principal investigator of the review, developed this second framework that classifies those reasons according to the following five main aspects of a clinical trial: (a) the intervention, (b) the trial design and conduct, (c) the outcomes, (d) the process and outcome measures used, and (e) the health condition addressed by the intervention. A summary of published examples of these two frameworks and its content can be found in Fetters and Molina-Azorin (2020).

Frameworks have been instrumental in illuminating the numerous possibilities that qualitative research can bring to the task of comprehensively and meaningfully evaluating interventions, particularly in the case of intervention researchers unfamiliar with MMR or skeptical of qualitative research. As a complement to more generic MMR methodological publications and textbooks, these frameworks have also served as practical guidelines for the design and implementation of MMR intervention studies. However, as described in several methodological reviews, predominantly in the health sciences, published empirical research consistently exhibits significant flaws in the reporting of design and implementation. Lewin et al. (2009) reviewed studies using qualitative research alongside randomized trials of complex healthcare interventions published during 2001 and 2003 and found that nearly half of them failed to report the qualitative sampling and analysis methods adequately, failed to justify the inclusion of a qualitative component, and failed to demonstrate integration. In the previously cited review by O'Cathain et al. (2013), the authors found that researchers frequently failed to explicitly acknowledge the contribution of the qualitative component to the study design and its added value. Similar findings were observed in a methodological review of the use and reporting quality of MMR in school-based obesity interventions by Brown et al. (2015), who reported that less than half of the studies justified the use of MMR and provided an adequate description of the MMR design. The authors also noted that, while most of the studies demonstrated evidence of integration of the quantitative and qualitative components, the reporting of this evidence frequently lacked detail and only a few studies described how it occurred. More recently, Thiessen et al. (2022) reviewed studies that combined RCTs and qualitative research in the field of oncology and concluded that the qualitative purpose was frequently not stated explicitly, the timing of the qualitative component within the overall design was frequently not reported, several aspects of the qualitative procedures were frequently not mentioned, and the integration of the quantitative and qualitative components was generally moderate. The methodological reporting flaws identified in these reviews warrant close examination because they may prevent researchers from fully communicating the unique insights afforded by an MMR approach.

While the literature on MMR intervention designs has contributed significantly to the advancement of this area of research practice, nearly all of these publications have been developed within the health sciences. To our knowledge, the only existing guidance on MMR intervention designs for researchers in the educational and social sciences was published by Grissmer (2016), who developed a guide that demonstrates the value of this type of design in evaluating educational and social interventions. This author asserted that the growing demand for MMR RCTs is a natural consequence of the current inadequacy of theories predicting social and educational outcomes. Since factors influencing outcomes of this type can be quite diverse due to the variety of the contexts in which interventions are implemented, existing theories may overlook some of these factors. Therefore, further developing these theories requires a more detailed and contextualized understanding of the multiple processes that contribute to the outcomes. According to Grissmer (2016), MMR intervention designs can contribute significantly to this understanding by clarifying the effects of context on intervention outcomes, elucidating why and how intervention effects occur, and explaining under what conditions the quantitative results are more reliable. Additionally, by generating this understanding, MMR designs of this type can be instrumental in elucidating the causal mechanisms underlying the long-term effects of the intervention (i.e., during a period after it is finalized).

The potential of MMR for generating contextualized knowledge is particularly relevant in the field of STEM (Science, Technology, Engineering, and Mathematics)-related interventions, as existing reviews indicate that a variety of contextual factors may contribute to differences in STEM education participation. For instance, van den Hurk et al. (2019) identified a number of factors that mediate and moderate participation in STEM education, some of which are context-specific, namely, the social context (i.e., educational policy, labor market/economy, and cultural environment/social views), the social environment (i.e., family and peers), and the school context (i.e., teaching pedagogy, school climate, and organization). Understanding these factors is crucial to developing successful interventions that would contribute to increasing interest in STEM programs and courses. Considering the limitations of quantitative research in properly capturing context, MMR intervention designs may enable researchers to achieve a more fine-grained and complete assessment of the range of contextual factors affecting the intervention outcomes. Additionally, this type of design can aid in the investigation of the long-term effects of STEM-related interventions, a subject that is particularly challenging to investigate due to the complexity of factors that act as long-term barriers to people becoming engaged in STEM (Prieto-Rodriguez et al., 2020).

The usefulness of MMR in evaluating STEM-related interventions was confirmed in a recent systematic review by Prieto-Rodriguez et al. (2020) of secondary school STEM interventions targeting girls. MMR was used in 19 of the 32 studies identified in that study. Despite the confirmed prevalence of these MMR studies, no reviews have been published that have systematically examined them. This omission is striking given the benefits of MMR intervention designs in developing context-specific knowledge that is easily transferable to policy and practice. Thus, an examination of the methodological features and reporting practices associated with this type of design is necessary to ascertain whether the added value of MMR is being realized in STEM-related interventions and whether the methodological limitations associated with MMR intervention designs in the health sciences also exist in this field. To address this need, our review aims to (1) characterize and describe the methodological features of MMR intervention studies intended to promote young people's interest in STEM; and (2) to assess the reporting quality of these studies. In this review, we intend to contribute to the practice of STEM intervention research by describing how MMR can improve the effective and comprehensive evaluation of STEM interventions and by making recommendations for reporting MMR intervention studies in this field.

2. Methods

2.1. Design

A methodological review was carried out. According to Mbuagbaw et al. (2020), methodological reviews are studies that report "on the design, conduct, analysis, or reporting of primary or secondary research-related reports" (p. 1). By examining the methodological characteristics of a sample of studies within a field identified systematically, reviews of this type can assist researchers in expanding their methodological repertoire, identifying existing methodological gaps, and improving future research practice (Aguinis et al., 2020; Martin et al., 2020; Howell Smith and Shanahan Bazis, 2021). Methodological reviews are particularly important in MMR intervention research because some basic procedures of the methodology are still not being implemented properly, as revealed by a number of reviews (see Section "Introduction"). The studies included in this methodological review were identified through a scoping review of intervention studies aimed at increasing young people's interest in STEM (Sáinz et al., 2022)—hereinafter referred to as the *original review*. Specifically, we focus here on the subsample of studies from the original review that used MMR. This review has been conducted and reported using the updated 2020 version of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). Since methodological reviews differ from conventional systematic reviews in their primary purpose and some of their procedures (Martin et al., 2020), only the PRISMA reporting criteria applicable to these reviews were used. Similarly, the protocol for this review was not registered due to the methodological nature of this study.

2.2. Eligibility criteria

To be included in the original review, publications had to: (a) report intervention studies aimed at increasing secondary school (i.e., in grades six and above according to the US educational system) students' participation in STEM fields; (b) clearly describe the intervention's objectives, participants, and context, as well as provide a concise description of its implementation; (c) evaluate the effectiveness of the intervention using a quantitative, qualitative, or MMR approach; and (d) be in English and published between 1998 and 2019. In the original review, non-empirical papers were excluded, including systematic reviews, editorials, and commentaries. In the methodological review, the same inclusion criteria as the original review were followed, except for the publication type, which was limited to peer-reviewed journal articles. In addition, studies included in the methodological review had to: (a) report quantitative research measuring outcomes specific to increasing secondary school students' interest in STEM fields using a pre- post-measurement; (b) report qualitative research carried out before, during, or after the quantitative component; and (c) provide evidence of integration of the qualitative and quantitative components; include a description of where and how the integration was carried out; refer to an attempt at integrating methods, or else use words associated with integration.

2.3. Information sources and search strategy

In the original review, we searched the title and abstract of publications in English indexed between 1998 and 2019 in the following five databases: APA PsycNET, ERIC, ProQuest, Scopus, and Web of Science. The searches in all databases were carried out on February 5, 2020. We used search terms related to the following four concept areas: intervention (e.g., program*, interven*, course*), STEM studies and professions (e.g., STEM, math*, science*), outcomes (e.g., interest*, engag*, motivat*), and gender (e.g., gender, girl*, female*) (see **Supplementary File 1** for the complete search query). The search strategy was developed in collaboration with an information scientist from the Universitat Oberta de Catalunya. In addition, we used three complementary search strategies to uncover relevant literature that database searches were unable to locate. First, we hand searched the following 20 journals publishing educational and behavioral STEM-related interventions: *American Psychologist*, *Annual Review of Psychology*, *Developmental Psychology*, *Educational Psychology Review*, *Educational Research, International Journal of Science Education*, *Journal of Applied Developmental Psychology*, *Journal of Educational Psychology*, *Journal of Experimental Child Psychology*, *Journal of Personality and Social Psychology*, *Personality and Social Psychology Bulletin*, *Perspectives on Psychological Science*, *Psychological Bulletin*, *Psychological Science*, *Psychology of Women Quarterly*, *Review of Educational Research, Science*, *Sex Roles*, *Social Psychological and Personality Science*, and *Social Science Quarterly*. Second, we reviewed the lists of publications of important authors in the field. Third, we scanned the references sections of key articles.

2.4. Selection process

The study selection was carried out in two phases. In the screening phase, two researchers independently screened the titles and abstracts of a random sample of 10% of the publications. Disagreements between the two reviewers were resolved through discussion with the involvement of a third reviewer when necessary. The remainder of the publications were divided between the two reviewers. In the eligibility phase, the two reviewers independently assessed their full texts and documented the reasons for exclusion. Disagreements in this phase were again resolved by consensus. EPPI-Reviewer was used in this phase for abstract and full text screening.

2.5. Data collection process and synthesis methods

We extracted and synthesized data from the studies included in this methodological review in three phases using qualitative content analysis (Schreier, 2012). In Phase 1, we read the full sample of the included studies to familiarize ourselves with the literature base we would synthesize. In Phase 2, we used the insights gathered during the familiarization phase to revise and update the extraction form we would use in the review. The extraction form, which the first author had previously used in two methodological reviews (Fàbregues et al., 2020, 2022), was guided by the literature on MMR intervention studies, the Good Reporting of Mixed Methods Studies (GRAMMS) guidelines (O'Cathain et al., 2008), and Fetters et al.'s (2013) typology of integration approaches (see **Supplementary File 2** for the extraction form). In Phase 3, the first author used the extraction form from Phase 2 to extract passages from the included articles in Microsoft Excel. Data extraction was double-checked by three reviewers, and any disagreements were resolved by consensus. Phase 3 involved reviewing all the extracted passages and comparing them to identify patterns of similarity and differences in the methodological features of the articles. Literature summary tables were used in this phase (Younas and Ali, 2021). Additionally, we used Stapley et al.'s (2021) ideal-type analysis method to create a typology of the contributions of the qualitative component to generate evidence of effectiveness. Following the steps recommended by these authors, we analyzed the previously extracted passages reporting the rationale and insights for using an MMR approach and those providing evidence from the integration of the quantitative and qualitative components. These steps included the following: (a) familiarizing ourselves with the extracted passages from each of the included studies; (b) preparing a summary of these passages; (c) systematically comparing these summaries to form clusters (called "ideal types") of similar studies based on the contribution of the qualitative component to the overall MMR design; (d) generating descriptions of the resulting ideal types and identifying studies that best represented each type; and (e) assessing the credibility of the typology by requesting an independent researcher to reclassify the studies into their ideal types, using the previously developed ideal-type descriptions. Steps b–d were carried out using MAXQDA version 2022.

3. Findings

The database and complementary searches yielded 40,170 records after removing duplicates. Two hundred fifteen studies were identified after assessing eligibility. Of these, 34 met the inclusion criteria for this review (see **Supplementary File 3** for a list of included studies). The PRISMA flowchart of the review process is in **Figure 1**, along with the reasons for excluding publications in the eligibility phase.

3.1. General characteristics of the included studies

The complete characteristics of the 34 MMR intervention studies included in the review are shown in **Supplementary File 4**. More than three quarters (n = 28) of the studies were published in the last 10 years, of which 19 were published in 2016–2020 and nine in 2011–2015. Six studies were published before 2011, equally distributed in the periods 2000–2005 (n = 3) and 2006–2010 (n = 3). The studies were published in general (n = 4) and field-specific (n = 18) education journals, including those devoted to science education (n = 7), educational technology (n = 6), and other subfields (n = 5). Only four studies were published in non-educational journals. In 23 studies, the intervention took place in the United States,

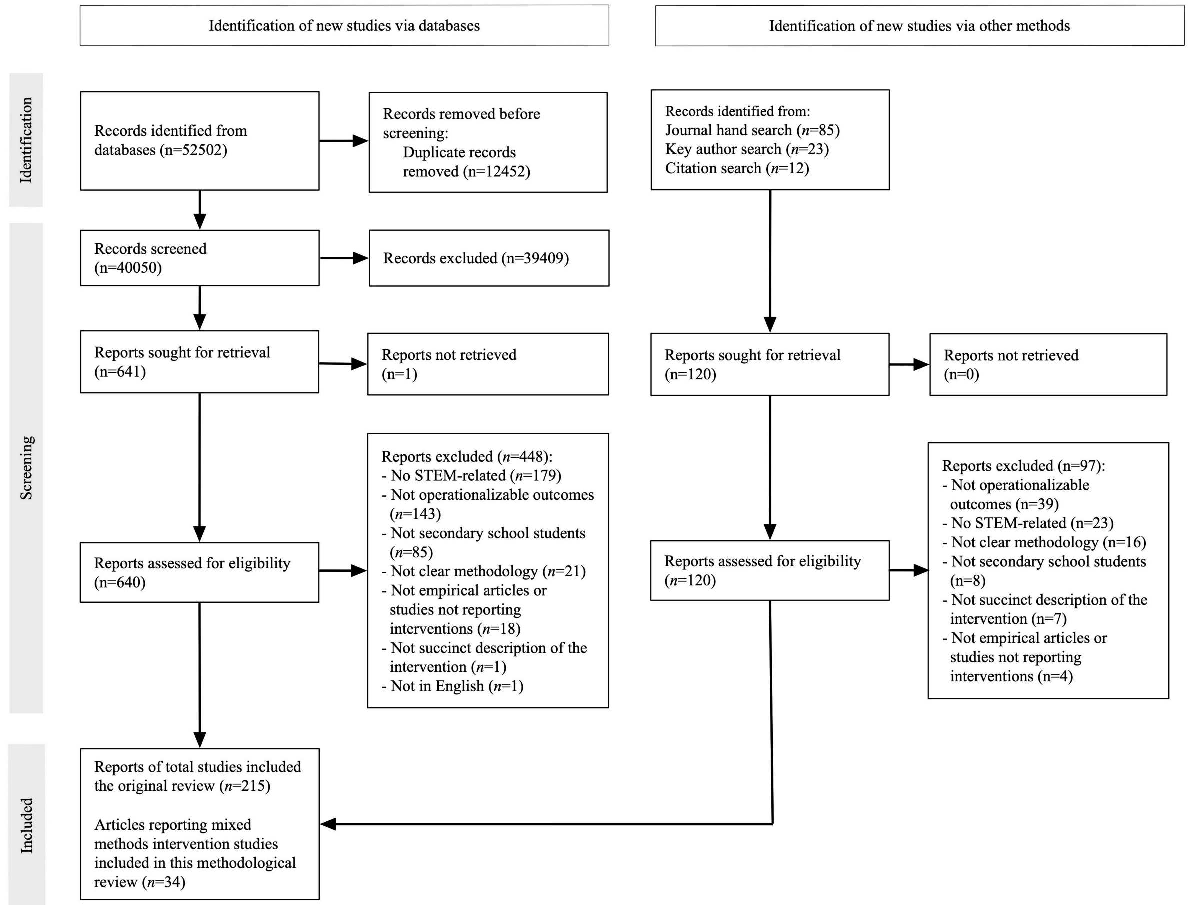

FIGURE 1
Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flowchart.

followed by three in the United Kingdom, two in Israel, and one in several countries, including Australia, Austria, Bolivia, Greece, Panama, South Africa, Spain, and Taiwan. Most of the interventions aimed to increase participants' interest in STEM (n = 12) and science (n = 11) fields, whereas fewer aimed to increase their interest in technology (n = 8) and STEAM (n = 3) fields. Motivation was one of the outcome measures in nearly all studies (n = 33), while achievement was measured in more than half of the studies (n = 16). Gender stereotypes (n = 8), identity (n = 6), emotional outcomes (n = 2), and academic choices (n = 1) were also addressed to a lesser extent. In 20 studies, students were both girls and boys, whereas, in 14 studies, participants were solely girls. More than half of the studies (n = 19) made explicit reference to a theoretical framework, either from psychology (e.g., expectancy value theory of motivation, social learning theory) or education (e.g., constructivist and learning-related theories).

Consistent with our inclusion criteria, all the included studies (n = 34) used the quantitative component to measure intervention effectiveness outcomes, while two of them also used this component to assess the acceptability of the intervention. Qualitative methods were also employed in the full sample of 34 studies to determine the perceived effectiveness of the intervention. Several studies replicated the pre-post quantitative data collection procedures in the qualitative component to assess changes in participants' views of intervention outcomes. For instance, Hughes et al. (2013) included an open-ended question on pre- and post-surveys to compare participants' levels of interest in STEM, self-concept related to STEM, and their perceptions of scientists before and after the intervention. In other cases, participants' views on the intervention effects were assessed retrospectively. For example, Aguilera and Perales-Palacios (2020) utilized a semi-structured interview with the participant teacher at the end of the intervention to elicit his views "on the effects of the intervention on student attitudes toward science and academic performance." Likewise, Archer et al. (2014) carried out focus groups with female students at the end of the intervention to explore their perceptions of whether "they had learnt anything" and if "they felt their attitudes to STEM careers had changed." In 11 studies, qualitative methods were used to evaluate the acceptability of the intervention, including "if they [the participants] had enjoyed participating in the [intervention]" (Ferreira, 2002), "which of the activities they liked and disliked" (Fabian and Topping, 2019), and "suggestions for improving [the intervention]" (Marino et al., 2013). Feasibility and fidelity were two other intervention domains examined in the qualitative component, each in three studies.

Half of the studies (*n* = 17) used the term "mixed methods" to describe the type of methods used, while the other half did not use any term. Only nine studies cited a methodological publication on MMR to justify this approach or explain its procedures. Five textbooks by Prof. John W. Creswell were among the six most cited publications, followed by the works of other influential MMR authors, such as Alan Bryman, Jennifer Greene, and Sharlene Hesse-Biber (see **Supplementary File 5** for a list of the key MMR publications cited by the included studies). None of these nine studies citing MMR methodological publications cited a publication explicitly focused on MMR intervention designs.

3.2. Methodological characteristics and reporting quality

Table 1 illustrates the reporting quality of the 34 studies in terms of their compliance with each of the six GRAMMS guidelines. **Supplementary File 6** gives further information regarding the methodological aspects of the studies.

TABLE 1 Reporting quality of the included studies in the review based on an adapted version of the good reporting of mixed methods studies (GRAMMS) guidelines (*n* = 34).

Guideline	Yes[a]	Yes, but[a]	No[a]
Describes the justification for using MMR to the research question	19	15	0
Describes the MMR design in terms of the purpose, priority, and sequence of methods	4	29	1
Describes each method in terms of sampling, data collection, and analysis	19	15	0
Reports evidence of integration[b]	30	1	3
Describes any limitation of one method associated with the presence of the other method	0	0	34
Describes any insights gained from mixing or integrating methods	6	3	25

MMR, mixed methods research.
[a] These categories are described in detail in **Supplementary File 7**.
[b] For the purpose of this study, the authors reformulated the original guideline number 4.

3.2.1. Rationale for using MMR

Despite the advantages of MMR over mono-method research for achieving additional insights into the studied phenomenon, it may not always be the appropriate option for addressing particular types of evaluations. Certain research questions may be better addressed using a quantitative or qualitative approach alone. For this reason, researchers must present a persuasive case for why MMR is the best approach for carrying out a particular intervention study. All 34 studies included in the review provided either an explicit or implicit rationale for choosing an MMR design to carry out the intervention study. This rationale was articulated explicitly in 19 studies, with most of these citing or quoting MMR methodological references to support the use of this methodology. Some of these studies reported rationales commonly cited in MMR textbooks, such as taking "advantage of the virtues of the quantitative and qualitative methodologies, compensating the weaknesses of one with the strengths of the other" (Aguilera and Perales-Palacios, 2020) or "bringing to light as many aspects as possible of students' activities in class" (Barak and Asad, 2012). In the remaining 15 studies, although this rationale was not explicitly stated, both quantitative and qualitative objectives were described in detail, allowing the reader to infer why an attempt was made to integrate both methodologies. Regardless of whether the rationales were explicitly or implicitly stated, in all the included studies, they were strongly tied to the ways in which the qualitative component complemented, strengthened, or supported the generation of evidence of effectiveness by the quantitative component. Examining these rationales, together with the integration outcomes and the insights gained from the use of MMR described in each article, we developed a typology of rationales for using qualitative research to generate additional evidence of effectiveness within the MMR study. **Table 2** shows these rationales, along with a description and an example for each. As shown in the table, the studies in our sample most frequently used the qualitative component to provide confidence in the integrity of the quantitative outcomes (*n* = 18); to enhance, augment, explain, or illustrate the quantitative outcomes (*n* = 14), to assist in identifying intervention components that may have influenced the quantitative outcomes (*n* = 8), and to help explain heterogeneity within the participants' responses to outcomes (*n* = 6).

3.2.2. MMR design

Mixed methods research studies should report the elements of their procedural design, including the sequencing of the quantitative and qualitative components (i.e., the timing of their execution) and whether one had priority over the other. Several typologies of MMR designs have been published, the most well-known of these developed and refined over the past 20 years by Creswell et al. (2003) and

Creswell and Plano Clark, 2007, 2011, 2018. Only four of the 34 studies included in this review provided a detailed explanation of the MMR design employed (Hur et al., 2017; Broder et al., 2019; Aguilera and Perales-Palacios, 2020; Chapman et al., 2020). All four of these cited one of Creswell's typologies of MMR designs to support the assertion that they used a convergent design. Convergent designs involve the separate collection of quantitative and qualitative data, followed by their integration for comparison or combination. In accordance with this approach, Hur et al. (2017) collected two distinct databases (i.e., quantitative data using surveys and qualitative data using focus groups, participant observation, and open-ended questions), each tentatively having equal priority, and integrated them during the analysis and interpretation phases to enhance the trustworthiness of the study. In one of these four studies (Broder et al., 2019), the design was incorrectly labeled. While the authors claimed to have used a sequential design, they actually employed a convergent design because one database did not inform the other, as is the case with sequential designs. Twenty-nine studies did not specify the type of MMR design used, but they did describe the sequence of the components, namely the time at which qualitative approaches were utilized within the MMR intervention design. Only one study did not indicate the MMR design type as well as the sequencing and priority of the components.

3.2.3. Quantitative and qualitative components

In addition to the specific MMR features, an MMR study must include quantitative and qualitative components

TABLE 2 Rationales for using qualitative research to generating additional evidence of effectiveness.

Rationale	Description	Example
Corroborate the QUAN findings ($n = 18$)	QUAL provides confidence about the integrity of the QUAN outcomes	"The interviews (. . .) reinforce the trends we detected in the quantitative assessment; student quotes revealed positive impacts of the program on scientific self-efficacy, interest in pursuing STEM in the future, as well as the importance of dissemination in shaping their identity as a scientist" (Broder et al., 2019)
Determine why and how the outcomes occurred ($n = 14$)	QUAL findings are used to enhance, augment, explain, or illustrate the QUAN outcomes	"Table 3 above shows that both career interest ($Z = 4.70$, $p < 0.001$) and intrinsic interest ($Z = 3.41$, $p = 0.001$) reported significantly higher scores on the post-test after our 1-week camp (. . .). In the interviews, all campers mentioned at least one reason why the camp contributed to their interest in programming" (Clarke-Midura et al., 2019)
Identify intervention strengths and weaknesses ($n = 8$)	QUAL findings assist in identifying intervention components that may have affected the QUAN outcomes	"Over the course of the day, the girls also came to see science as more interesting and enjoyable (i.e., as having higher intrinsic value); qualitative data indicated that this was due to the variety of topics covered and the hands-on activities in which they participated" (Skipper and de Carvalho, 2019)
Explain differences in effectiveness within the sample ($n = 6$)	QUAL helps explain or better identify variability in participant responses to QUAN outcomes	"Male students consistently rated the activities higher than the girls. Their VMT scores were also significantly higher. In the student interviews, female students provided fewer positive responses about the intervention than male students. One possible reason for this is the nature of the paired work. Some girls that were paired with boys did not manage to work particularly well with their partner as the boys tended to take control of the tablet. This hesitation to work with the opposite sex was mentioned several times in the interviews" (Fabian and Topping, 2019)
Identify additional intervention benefits ($n = 2$)	QUAL helps identify additional benefits in addition to those represented by positive QUAN outcomes	"In addition to our two primary research questions, we also explored any additional benefits for students from participating in EPICC. The follow-up [qualitative] surveys pointed to a number of lasting impacts on participants. For instance, students reported that their experience in the service-learning project helped them feel a sense of contribution and connectedness to other people, as well as gratitude" (Collins et al., 2020)
Overcome study weaknesses by utilizing multiple sources of evidence ($n = 1$)	QUAL is added to other sources of QUAN data to compensate for the study's inherent limitations (e.g., small sample of intervention participants)	"(. . .) the small number of students who participated in this exploratory work prohibits generalization. Nevertheless, questionnaire and interview data, personal observations both from teachers and researcher (. . .) and the development of text quality all indicate that the concept presented here may contribute to a positive interest development amongst high-school students with respect to NaSc" (Simon et al., 2016)
Improve confidence in the use of QUAN measures ($n = 1$)	QUAL and QUAN data are triangulated to obtain confidence in the application of QUAN measures when their validity and reliability have not yet been tested	"(. . .) the middle school surveys have not been administered to enough participants to be declared reliable and valid—this process is currently occurring. As a result, we chose to do our own reliability tests and use the quantitative data as a source of triangulation for the qualitative data (. . .) The qualitative codes matched the quantitative categories in 90% of the instances. With this triangulation using our qualitative data, we were confident in our decision to use these measures" (Hughes et al., 2013)
Reveal conflicting findings ($n = 1$)	QUAL findings conflict with QUAN outcomes, demonstrating the need for further inquiry	"This construct [understanding of computing jobs] was measured using two sources of data to determine whether students' understanding changed over time, and they revealed conflicting findings" (Denner et al., 2012)

QUAN, quantitative; QUAL, qualitative.

that are elaborated with technical competence and reported transparently. While 19 of the studies reported in detail all quantitative and qualitative procedures, including sampling, data collection, and analysis, 15 studies failed to accurately report at least two of these procedures. In those studies, the authors frequently did not describe the methods used to analyze the qualitative data (n = 10) and/or the criteria that informed the selection of the qualitative sample (n = 8). For instance, while some of them described the characteristics of the participants in the qualitative component, they did not indicate why and how the researchers selected that particular group of participants over others.

In the quantitative component, single-group pre- and post-treatment designs (n = 25) were employed the most often, while other types of designs, such as multiple-group pre- and post-treatment designs (n = 5) and RCTs (n = 4), were employed much less frequently. In the qualitative component, only one study reported the qualitative design used, and this was ethnography. In all the studies, questionnaires were the primary quantitative data collection method (n = 34), accompanied in some cases by achievement exams (n = 5), content knowledge tests (n = 3), observation checklists (n = 2), quantitative content analysis (n = 2), and other methods (n = 3). In the qualitative component, the methods used were interviews (n = 23), open-ended questions (n = 14), focus groups (n = 11), observations (n = 8), and other methods (n = 5). The use of multiple data collection methods was marginally less prevalent in the quantitative component (n = 14) than in the qualitative component (n = 16). Lastly, qualitative data were obtained at various different times throughout each study, specifically in 12 studies before the intervention, in 10 studies during the intervention, in 30 studies immediately after finishing the intervention, and in six studies a few months after the intervention was completed. In only one study this information was not clear.

3.2.4. Integration

In an MMR intervention study, integration involves mixing quantitative and qualitative components in one or more phases of a study to generate insights that lead to a more precise and exhaustive evaluation of the intervention. To effectively communicate these insights, researchers must provide a precise description of the integration outcomes and the resulting knowledge. Thirty of the 34 studies included in our review provided explicit evidence of integration; three did not provide any evidence; and in one study, the insights gained from integration could be inferred.

In the studies that provided explicit or partial integration reporting, we coded how integration was carried out using Fetters et al.'s (2013) typology of integration strategies. These authors explained that integration can occur through merging (when the two types of data or findings are brought together for comparison or analysis), building (when the findings from one component are used to define the data collection strategy of the other component) and connecting (when the findings from one component are used to define the sampling strategy of the other component). Thirty-one studies integrated through merging, two studies integrated through building and only one integrated through connecting. When merging was employed, the authors described the relationship between the quantitative and qualitative findings, including whether one form of data confirmed, expanded, or contradicted the findings of the other type. For instance, in a study evaluating two informal science programs, Hughes et al. (2013) described how the quantitative t-test findings indicating positive changes in the participants' STEM identity confirmed the interview findings, which "also provided qualitative evidence of their [the participants'] improved trajectories." Conversely, in an evaluation study of a project-service learning curriculum for high school students, Ruth et al. (2019) explained how the quantitative findings contradicted the qualitative ones. While, according to the quantitative findings, the project under evaluation was "not creating much change in the skills domains that could support any students', including URM (historically underrepresented minority) and female students', pathways into Engineering/STEM," the qualitative ethnographic data indicated it was "positively impacting URM and female students in particular, and in ways that are meaningful and could potentially orient them toward STEM."

In the two studies that integrated through building, the authors used the findings from one component to inform the data collection approach of the other component. Based on the quantitative data, Magerko et al. (2016) concluded that the learning module EarSketch was effective in enhancing students' computing content knowledge and intent to persist in computing. To fully understand the success of this module, the authors carried out two focus groups using an interview guide based on the main conclusions from the quantitative findings. Lastly, in the study by Hughes et al. (2013) cited above, the authors integrated through connecting by selecting the participants in the interviews based on their scores in the quantitative measures (e.g., STEM self-concept, parental education, and exposure to STEM role models) to build a heterogeneous qualitative sample.

Using the same typology described by Fetters et al. (2013), we classified the ways in which integration was reported. All studies that provided explicit or implicit evidence of integration (n = 31) used a narrative approach to report the relationship between the two types of data. In these studies, this relationship was frequently explained verbally in both the results and discussion sections (n = 16), and less frequently in the results (n = 8) or discussion (n = 7) sections alone. Overall, the authors devoted substantial space to elucidating the interrelationships between the different quantitative and qualitative findings, thereby contributing to a more robust reporting of the integration outcomes. No studies used tables, diagrams, matrices, or figures to visually integrate the findings in the form of joint displays.

3.2.5. Limitations and insights

No limitations as a result of using one methodological approach in conjunction with the other were reported in any of the articles. Furthermore, only six publications offered a clear description of the added value gained by utilizing an MMR approach in the discussion or conclusions sections. In these studies, authors declared that MMR allowed them to "gain invaluable insights on the effects of the games that could not have been discovered only through quantitative tests" (Kebritchi et al., 2010), or provide "different levels of granularity in the investigation of the effects" of the intervention (Fabian and Topping, 2019), amongst others. In three studies, the added value was not explicitly stated, but could be inferred.

4. Discussion

4.1. Summary of main findings

This review is, to the best of our knowledge, the first work to examine the use of MMR intervention designs in education and the social sciences. It is of particular interest given that most of the methodological publications about MMR intervention designs deal with the health sciences. As well as providing guidance for implementing designs of this type, these publications have shown that MMR intervention designs are becoming more popular owing to their usefulness in expanding the scope and strengthening the credibility of intervention evaluations in the health sciences. In this review, we examine whether such prevalence and advantages are also present in MMR intervention studies in the social and educational sciences in light of recent claims that such designs provide essential context and population-specific information for these interventions (Grissmer, 2016).

Our findings show an increase in the publication of MMR studies of interventions aimed at stimulating young people's interest in STEM, with more than half of these studies having been published since 2016. This conclusion is congruent with the findings of a recent review of MMR interventions in emotional and behavioral disorders by Fàbregues et al. (2022), which identified a similar increase in the number of MMR intervention studies in that field. Moreover, our findings show that the incorporation of qualitative approaches into quantitative intervention designs was especially helpful in the study of interventions aimed at enhancing young people's engagement in STEM, and particularly in elucidating how, under what conditions, and for what types of populations such interventions were successful or unsuccessful. For instance, in a study analyzing the success of a computer science programming summer camp for middle school kids, Clarke-Midura et al. (2019) used interviews with camp attendees to gain valuable insights on why and how the positive quantitative outcomes occurred. In qualitative interviews, the authors were able to discover numerous social elements that influenced participants' positive shift in interest in STEM, including the opportunity to show their parents the apps they had built, the ability to provide and receive advice, and/or the availability of mentors. In another study assessing the impact of mobile technology on success in mathematics, Fabian and Topping (2019) were able to qualitatively discover that the intervention effects varied by gender, a conclusion that could not have been reached using purely quantitative methods. The authors determined, through student interviews, that male students viewed the activities more favorably than female students because some female students were matched with males who frequently assumed full control of the tablet. Furthermore, in several studies, a qualitative component was included for triangulation purposes to bolster the quantitative findings. In addition to quantitative measures of self-efficacy and interest in STEM subjects, Broder et al. (2019) used data from qualitative interviews to confirm the beneficial patterns revealed in the quantitative component. These trends suggest that the authors of the included studies were aware of the benefits of MMR intervention designs and employed them for the reasons cited in the health sciences frameworks described above. However, none of the included studies cited these frameworks or any methodological work on the combination of qualitative research and quantitative trials. This conclusion is relevant because it implies that the use of MMR in intervention designs was driven more by the intention to answer specific research objectives than by the literature.

In terms of reporting the MMR components, the 34 studies included in our study displayed a generally high level of quality. All the studies provided a justification for using an MMR intervention design, and more than half of them did so explicitly. This finding contrasts with previous reviews of MMR intervention designs in the health and behavioral sciences (Lewin et al., 2009; O'Cathain et al., 2013; Broder et al., 2019), in which the rationale for incorporating a qualitative component into a quantitative intervention design was either not provided or not detailed enough. Clarity in the reporting of the justification for using MMR was enhanced by the fact that, in most studies, the reason for including a qualitative component was explicitly mentioned, allowing the reader to understand how the qualitative aim interacted with the quantitative purpose. Integration of the quantitative and qualitative elements was also well-reported. Nearly all the studies integrated by merging, and the majority of these clearly reported the integration outcomes. Often, the reporting of integration was enhanced by extensively describing the outcomes in several subsections of the studies, particularly in both the findings and discussion sections. This form of reporting is consistent with Bazeley's

(2015) suggestion that integration reporting should not be restricted to the discussion section in order to maximize the integrative potential of MMR. According to this author, a greater emphasis on making explicit the linkages between the findings of both components throughout the entire manuscript, as was the case in several of the studies reviewed, could certainly result in a better integration. In addition, our review findings contradict the results of earlier reviews of MMR intervention designs, in which studies modestly reported integration. For instance, in a recent review of these types of designs in oncology, Thiessen et al. (2022) concluded that integration was often "limited to a brief statement regarding how a study conclusion was supported by both the qualitative and RCT data." Not often was this the case among the studies analyzed in this review.

However, despite the generally good reporting quality of the reviewed studies, we nevertheless found three main issues. First, very few studies provided an accurate description of the design used, even though in most of them it was possible to identify the relative timing of the qualitative and quantitative components, a finding that contrasts with the findings from O'Cathain et al. (2013) and Thiessen et al.'s (2022) reviews of qualitative research utilized with RCTs in the health sciences. Second, even though the reporting of the methods followed in the quantitative component was generally detailed, the description of several qualitative procedures, particularly the qualitative sampling and analysis, lacked the same level of detail. Similar reporting issues with these two qualitative aspects have been identified in the reviews by Lewin et al. (2009) and Fàbregues et al. (2022). Third, no studies integrated both types of data using joint displays, which are visual tools based on tables or figures for performing and representing integration in MMR more clearly (Guetterman et al., 2015, 2021). Previous methodological reviews of MMR intervention designs have also found that none or very few of the included studies used displays of this type (Fàbregues et al., 2022).

4.2. Implications for the reporting of MMR studies for evaluating STEM-related interventions

Based on the previously observed inconsistencies in the reporting of MMR in the field of STEM-related interventions, we can draw three implications for authors of studies of this type. First, authors must describe the type of MMR design used in their studies, either by citing one of the existing typologies of MMR designs or by providing details of the purpose of their design, the timing of the quantitative and qualitative components, and the points of mixing between these components throughout the study. This can be achieved, for instance, through the use of procedural diagrams. According to Creswell and Plano Clark (2018), procedural diagrams can facilitate the intuitive representation of the MMR study features, thereby making it easier for the readers to "convey the complexity of mixed methods designs." This is particularly relevant for MMR intervention studies due to the greater degree of complexity than other MMR studies using core designs (i.e., convergent, explanatory sequential, and exploratory sequential designs). Second, authors must provide transparent and accurate reporting of qualitative methods, including qualitative sampling and data analysis procedures. Even though the quantitative component tends to play a prominent role in these types of designs, this does not imply that the qualitative component should not adhere to adequate reporting standards. Third, in addition to presenting the integration findings in narrative format, authors should include joint displays that illustrate the researchers' "cognitive process of merging, comparing, relating, and linking qualitative and quantitative data or results" (Guetterman et al., 2021). If the authors of the included studies had used these types of displays, integration would have been represented more clearly, making it easier for the reader to identify the meta-inferences (i.e., inferences derived from the integration of quantitative and qualitative findings in the form of theoretical statements, narratives, or a story) resulting from the MMR study.

4.3. Strengths and limitations

This study is a follow-up to a larger review of intervention studies aimed at increasing the participation of young people in STEM. Since the original review included all types of studies and no MMR-specific terms were used, we were able to accurately identify all MMR studies, including those that did not use this term (i.e., a total of 17 studies). In other words, we did not need to use method-specific terms because the initial sample included all relevant studies, including those that utilized MMR. The study had also some limitations. First, authors may use a wide variety of terms to describe the topic of the intervention, making it difficult to locate these types of studies in systematic reviews. Consequently, due to the search terms employed, it is possible that we overlooked several pertinent studies. Second, because MMR is still a developing methodology and some of its reporting components require further operationalization (e.g., evidence of integration), it is likely that authors of the included studies will disagree with some of our decisions during the extraction and coding processes. Third, we limited our evaluation of the quality of the included studies to the quality of the reporting and not the methodological quality. Future reviews could evaluate components of this other dimension of quality, such as whether the quantitative and qualitative components adhered to the quality criteria of each tradition or whether the divergences between the quantitative and qualitative findings have been adequately addressed (Hong et al., 2018, 2019).

5. Conclusion

In recent years, MMR has been widely utilized in intervention studies aimed at fostering an interest in STEM among young people. In these studies, researchers have incorporated qualitative research to overcome significant limitations of quantitative intervention designs to provide contextual knowledge easily transferable to practice. The included studies were generally adequately reported, particularly regarding the justification for adopting MMR and the integration of quantitative and qualitative data, two crucial components of MMR. However, some room for improvement was observed in a few components, namely, the description of the type of MMR design used, the explanation of the procedures in the qualitative component, and the use of joint displays for the systematic and visual representation of integration. More attention to these reporting standards will help ensure that the potential of MMR to provide a more comprehensive evaluation of the intervention is clearly communicated to readers.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding authors.

Author contributions

SF and MS: review conceptualization and design. SF: search strategy and writing—original draft. SF, MR, and B-SL-P: screening and study eligibility. SF, MS, MR, EE-B, and B-SL-P: data extraction and synthesis. All authors writing—review and editing and approved the submitted version.

Funding

This study belonged to a research project led by MS and funded by the Spanish Ministry of Economy, Industry and Competitiveness (MINECO), Spanish State Research Agency (AEI), and European Regional Development Fund (ERDF) grant number (FEM2017-84589-R).

Acknowledgments

The authors would like to acknowledge the help of Dick Edelstein in editing 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

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.956300/full#supplementary-material

References

Aguilera, D., and Perales-Palacios, F. J. (2020). Learning biology and geology through a participative teaching approach: the effect on student attitudes towards science and academic performance. *J. Biol. Educ.* 54, 245–261. doi: 10.1080/00219266.2019.1569084

Aguinis, H., Ramani, R. S., and Alabduljader, N. (2020). Best-practice recommendations for producers, evaluators, and users of methodological literature reviews. *Organ. Res. Methods* 2020:1094428120943281. doi: 10.1177/1094428120943281

Archer, L., DeWitt, J., and Dillon, J. (2014). 'It didn't really change my opinion': Exploring what works, what doesn't and why in a school science, technology, engineering and mathematics careers intervention. *Res. Sci. Technol.* 32, 35–55. doi: 10.1080/02635143.2013.865601

Aschbrenner, K. A., Kruse, G., Gallo, J. J., and Plano Clark, V. L. (2022). Applying mixed methods to pilot feasibility studies to inform intervention trials. *Pilot Feasibil. Stud.* 8:217. doi: 10.1186/s40814-022-01178-x

Barak, M., and Asad, K. (2012). Teaching image-processing concepts in junior high school: Boys' and girls' achievements and attitudes towards technology. *Res. Sci. Tech. Educ.* 30, 81–105. doi: 10.1080/02635143.2012.656084

Bazeley, P. (2015). "Writing up multimethod and mixed methods research for diverse audiences," in *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*, eds S. Hesse-Biber and R. B. Johnson (New York, NY: Oxford University Press), 296–313.

Boeije, H. R., Drabble, S. J., and O'Cathain, A. (2015). Methodological challenges of mixed methods intervention evaluations. *Methodology* 11, 119–125.

Bouchard, K., and Tulloch, H. (2020). Strengthening behavioral clinical trials with online qualitative research methods. *J. Health Psychol.* 25, 256–265. doi: 10.1177/1359105319855120

Broder, E. D., Guilbert, K. E., Tinghitella, R. M., Murphy, S. M., Ghalambor, C. K., and Angeloni, L. M. (2019). Authentic science with dissemination increases self-efficacy of middle school students. *Integr. Comp. Biol.* 59, 1497–1508. doi: 10.1093/icb/icz140

Brown, K. M., Elliott, S. J., Leatherdale, S. T., and Robertson-Wilson, J. (2015). Searching for rigour in the reporting of mixed methods population health research: A methodological review. *Health Educ. Res.* 30, 811–839. doi: 10.1093/her/cyv046

Chapman, A., Rodriguez, F. D., Pena, C., Hinojosa, E., Morales, L., Del Bosque, V., et al. (2020). "Nothing is impossible": Characteristics of Hispanic females participating in an informal STEM setting. *Cult. Stud. Sci.* 15, 723–737. doi: 10.1007/s11422-019-09947-6

Cheng, K. K. F., and Metcalfe, A. (2018). Qualitative methods and process evaluation in clinical trials context: Where to head to?. *Int. J. Qual. Methods* 17:160940691877421. doi: 10.1177/1609406918774212

Clarke-Midura, J., Sun, C., Pantic, K., Poole, F., and Allan, V. (2019). Using informed design in informal computer science programs to increase youths' interest, self-efficacy, and perceptions of parental support. *ACM Trans. Comput. Educ.* 19, 1–24. doi: 10.1145/3319445

Collins, M. A., Totino, J., Hartry, A., Romero, V. F., Pedroso, R., and Nava, R. (2020). Service-learning as a lever to support STEM engagement for underrepresented youth. *J. Exp. Educ.* 43, 55–70. doi: 10.1177/1053825919887407

Creswell, J. W., and Plano Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage.

Creswell, J. W., and Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage.

Creswell, J. W., and Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., and Hanson, W. E. (2003). "Advanced mixed methods research designs," in *Handbook of Mixed Methods in Social and Behavioral Research*, eds A. Tashakkori and C. Teddlie (Thousand Oaks, CA: Sage), 209–238.

Denner, J., Werner, L., Martinez, J., and Bean, S. (2012). Computing goals, values, and expectations: Results from an after-school program for girls. *J. Women Minor. Sci. Eng.* 18, 199–213. doi: 10.1615/JWomenMinorScienEng.2013002075

Fabian, K., and Topping, K. J. (2019). Putting "mobile" into mathematics: Results of a randomised controlled trial. *Contemp. Educ. Psychol.* 59:101783. doi: 10.1016/j.cedpsych.2019.101783

Fàbregues, S., Hong, Q. N., Escalante-Barrios, E. L., Guetterman, T. C., Meneses, J., and Fetters, M. D. (2020). A methodological review of mixed methods research in palliative and end-of-life care (2014–2019). *Int. J. Environ. Res. Public Health* 17, 1–16. doi: 10.3390/ijerph17113853

Fàbregues, S., Mumbardó-Adam, C., Escalante-Barrios, E. L., Hong, Q. N., Edelstein, D., Vanderboll, K., et al. (2022). Mixed methods intervention studies in children and adolescents with emotional and behavioral disorders: A methodological review. *Res. Dev. Disabil.* 126:104239. doi: 10.1016/j.ridd.2022.104239

Ferreira, M. (2002). Ameliorating equity in science, mathematics, and engineering: A case study of an after-school science program. *Equity Excell. Educ.* 35, 43–49. doi: 10.1080/713845242

Fetters, M. D. (2020). *The Mixed Methods Research Workbook: Activities for Designing, Implementing, and Publishing Projects*. Thousand Oaks, CA: Sage.

Fetters, M. D., Curry, L. A., and Creswell, J. W. (2013). Achieving integration in mixed methods designs - Principles and practices. *Health Serv. Res.* 48, 2134–2156. doi: 10.1111/1475-6773.12117

Fetters, M. D., and Molina-Azorin, J. F. (2020). Utilizing a mixed methods approach for conducting interventional evaluations. *J. Mix. Methods Res.* 14, 131–144. doi: 10.1177/1558689820912856

Grissmer, D. W. (2016). *A Guide to Incorporating Multiple Methods in Randomized Controlled Trials to Assess Intervention Effects*. Washington, DC: American Psychological Association.

Guetterman, T. C., Fàbregues, S., and Sakakibara, R. (2021). Visuals in joint displays to represent integration in mixed methods research: A methodological review. *Methods Psychol.* 5:100080. doi: 10.1016/j.metip.2021.100080

Guetterman, T. C., Fetters, M. D., and Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *Ann. Fam. Med.* 13, 554–561. doi: 10.1370/afm.1865

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., et al. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Educ. Inf.* 34, 285–291. doi: 10.3233/EFI-180221

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., et al. (2019). Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. *J. Clin. Epidemiol.* 111, 49-59.e1. doi: 10.1016/j.jclinepi.2019.03.008

Howell Smith, M. C., and Shanahan Bazis, P. (2021). Conducting mixed methods research systematic methodological reviews: A review of practice and recommendations. *J. Mix. Methods Res.* 15, 546–566. doi: 10.1177/1558689820967626

Hughes, R. M., Nzekwe, B., and Molyneaux, K. J. (2013). The single sex debate for girls in science: A comparison between two informal science programs on middle school students' STEM identity formation. *Res. Sci. Educ.* 43, 1979–2007. doi: 10.1007/s11165-012-9345-7

Hur, J. W., Andrzejewski, C. E., and Marghitu, D. (2017). Girls and computer science: Experiences, perceptions, and career aspirations. *Comput. Sci. Educ.* 27, 100–120. doi: 10.1080/08993408.2017.1376385

Johnson, R. B., and Schoonenboom, J. (2016). Adding qualitative and mixed methods research to health intervention studies: Interacting with differences. *Qual. Health Res.* 26, 587–602. doi: 10.1177/1049732315617479

Kebritchi, M., Hirumi, A., and Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Comput. Educ.* 55, 427–443. doi: 10.1016/j.compedu.2010.02.007

Landes, S. J., McBain, S. A., and Curran, G. M. (2019). An introduction to effectiveness-implementation hybrid designs. *Psychiatry Res.* 280:112513. doi: 10.1016/j.psychres.2019.112513

Lewin, S., Glenton, C., and Oxman, A. D. (2009). Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: Methodological study. *BMJ* 339, 732–734. doi: 10.1136/bmj.b3496

Magerko, B., Freeman, J., McKlin, T., Reilly, M., Livingston, E., McCoid, S., et al. (2016). EarSketch: A STEAM-based approach for underrepresented populations in high school computer science education. *ACM Trans. Comput. Educ.* 16, 1–25. doi: 10.1145/2886418

Maher, L., and Neale, J. (2019). Adding quality to quantity in randomized controlled trials of addiction prevention and treatment: a new framework to facilitate the integration of qualitative research. *Addiction* 114, 2257–2266. doi: 10.1111/add.14777

Marino, M. T., Israel, M., Beecher, C. C., and Basham, J. D. (2013). Students' and Teachers' perceptions of using video games to enhance science instruction. *J. Sci. Educ. Technol.* 22, 667–680. doi: 10.1007/s10956-012-9421-9

Martin, G. P., Jenkins, D. A., Bull, L., Sisk, R., Lin, L., Hulme, W., et al. (2020). Toward a framework for the design, implementation, and reporting of methodology scoping reviews. *J. Clin. Epidemiol.* 127, 191–197. doi: 10.1016/j.jclinepi.2020.07.014

Mbuagbaw, L., Lawson, D. O., Puljak, L., Allison, D. B., and Thabane, L. (2020). A tutorial on methodological studies: The what, when, how and why. *BMC Med. Res. Methodol.* 20:226. doi: 10.1186/s12874-020-01107-7

O'Cathain, A. (2018). *Using Qualitative Research with Randomized Controlled Trials*. Oxford: Oxford University Press.

O'Cathain, A., Murphy, E., and Nicholl, J. (2008). The quality of mixed methods studies in health services research. *J. Health Serv. Res. Policy* 13, 92–98. doi: 10.1258/jhsrp.2007.007074

O'Cathain, A., Thomas, K. J., Drabble, S. J., Rudolph, A., and Hewison, J. (2013). What can qualitative research do for randomised controlled trials? A systematic mapping review. *BMJ Open* 3:e002889. doi: 10.1136/bmjopen-2013-002889

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372:n71. doi: 10.1136/bmj.n71

Prieto-Rodriguez, E., Sincock, K., and Blackmore, K. (2020). STEM initiatives matter: results from a systematic review of secondary school interventions for girls. *Int. J. Sci. Educ.* 42, 1144–1161. doi: 10.1080/09500693.2020.1749909

Richards, D. A., Bazeley, P., Borglin, G., Craig, P., Emsley, R., Frost, J., et al. (2019). Integrating quantitative and qualitative data and findings when undertaking randomised controlled trials. *BMJ Open* 9:e032081. doi: 10.1136/bmjopen-2019-032081

Ruth, A., Hackman, J., Brewis, A., Spence, T., Luchmun, R., Velez, J., et al. (2019). Engineering projects in community service (EPICS) in high schools: Subtle but potentially important student gains detected from human-centered curriculum design. *Educ. Sci.* 9:35. doi: 10.3390/educsci9010035

Sáinz, M., Fàbregues, S., Romano, M. J., and López, B. S. (2022). Interventions to increase young people's interest in STEM. A scoping review. *Front. Psychol.* 13:954996. doi: 10.3389/fpsyg.2022.954996

Sandelowski, M. (1996). Using qualitative methods in intervention studies. *Res. Nurs. Health* 19, 359–364. doi: 10.1002/(SICI)1098-240X(199608)19:4<359:AID-NUR9<3.0.CO;2-H

Schreier, M. (2012). *Qualitative Content Analysis in Practice*. London: Sage.

Simon, U. K., Steindl, H., Larcher, N., Kulac, H., and Hotter, A. (2016). Young science journalism: Writing popular scientific articles may contribute to an increase of high-school students' interest in the natural sciences. *Int. J. Sci. Educ.* 38, 814–841. doi: 10.1080/09500693.2016.1173260

Skipper, Y., and de Carvalho, E. (2019). "I have seen the opportunities that science brings": Encouraging girls to persist in science. *Educ. Forum* 83, 199–214. doi: 10.1080/00131725.2019.1576820

Stapley, E., O'Keeffe, S., and Midgley, N. (2021). Essentials of ideal-type analysis: A qualitative approach to constructing typologies.

Thiessen, M., Harris, D., Pinches, A., Vaska, M., Moules, N., Raffin Bouchal, S., et al. (2022). Qualitative studies conducted alongside randomized controlled trials in oncology: A scoping review of use and rigour of reporting. *Int. J. Nurs. Stud.* 128:104174. doi: 10.1016/j.ijnurstu.2022.104174

van den Hurk, A., Meelissen, M., and van Langen, A. (2019). Interventions in education to prevent STEM pipeline leakage. *Int. J. Sci. Educ.* 41, 150–164. doi: 10.1080/09500693.2018.1540897

Younas, A., and Ali, P. (2021). Five tips for developing useful literature summary tables for writing review articles. *Evid. Based Nurs.* 24, 32–34. doi: 10.1136/ebnurs-2021-103417

Zhang, W. (2014). Mixed methods application in health intervention research: A multiple case study. *Int. J. Mult. Res.* 8, 24–35. doi: 10.5172/mra.2014.8.1.24