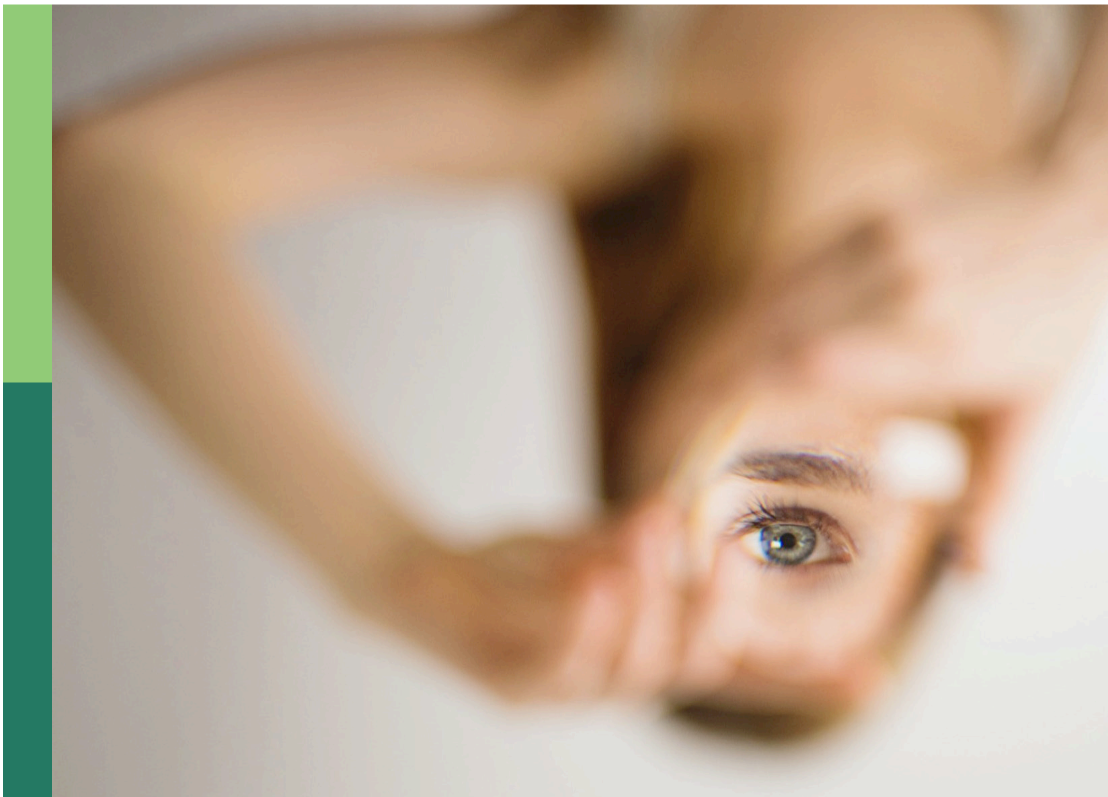
frontiers | Research Topics

Methods and applications in educational psychology

Edited by
Antonio P. Gutierrez de Blume, Kai-Yu Tang,
María Soledad Segretin, Christelle Declercq and Mei-Shiu Chiu

Published in
Frontiers in Psychology
Frontiers in Education



ISSN 1664-8714
ISBN 978-2-8325-2840-2
DOI 10.3389/978-2-8325-2840-2

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

Methods and applications in educational psychology

Topic editors
Antonio P. Gutierrez de Blume — Georgia Southern University, United States
Kai-Yu Tang — National Chung Hsing University, Taiwan
María Soledad Segretin — National Scientific and Technical Research Council (CONICET), Argentina
Christelle Declercq — Université de Reims Champagne-Ardenne, France
Mei-Shiu Chiu — National Chengchi University, Taiwan

Citation
Gutierrez de Blume, A. P., Tang, K.-Y., Segretin, M. S., Declercq, C., Chiu, M.-S., eds. (2023). *Methods and applications in educational psychology*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2840-2

Table of contents

05 **Editorial: Methods and applications in educational psychology**
María Soledad Segretin, Mei-Shiu Chiu, Kai-Yu Tang, Antonio P. Gutierrez de Blume and Christelle Declercq

08 **A Co-word Analysis of Selected Science Education Literature: Identifying Research Trends of Scaffolding in Two Decades (2000–2019)**
Tzu-Chiang Lin, Kai-Yu Tang, Shu-Sheng Lin, Miao-Li Changlai and Ying-Shao Hsu

22 **Development and Testing of the Curiosity in Classrooms Framework and Coding Protocol**
Jamie J. Jirout, Sharon Zumbrunn, Natalie S. Evans and Virginia E. Vitiello

36 **The Effect of Physical Exercise on Depression in College Students: The Chain Mediating Role of Self-Concept and Social Support**
Junliang Zhang, Shuang Zheng and Zhongzheng Hu

42 **Identification of Cognitive Activities That Underlie Variations in Lecture Note-Taking: An Exploration of Japanese and Chinese High School Students' Strategies in Mathematics Class**
Mengsi Liu and Yuri Uesaka

54 **Leveraging a multidimensional linguistic analysis of constructed responses produced by college readers**
Joseph P. Magliano, Lauren Flynn, Daniel P. Feller, Kathryn S. McCarthy, Danielle S. McNamara and Laura Allen

68 **Systematic literature review on audio-visual multimodal input in listening comprehension**
Tan Shaojie, Arshad Abd Samad and Lilliati Ismail

79 **A phenomenological study of online assessment during a pandemic crisis: Insights from Malaysia, Lithuania, and Spain**
Nagaletchimee Annamalai, Antonia Ramírez García, Viktorija Mažeikienė, Marwan Harb Alqaryouti, Radzuwan Ab Rashid and Arulselvi Uthayakumaran

90 **Analysis of the effect of cognitive ability on academic achievement: Moderating role of self-monitoring**
Yueqi Shi and Shaowei Qu

104 **Translation and validation of the high-school satisfaction scale (H-SatP Scale) in Peruvian students**
Johana Gonzales-Valdivia, Wilter C. Morales-García, Jacksaint Saintila, Salomón Huancahuire-Vega, Mardel Morales-García and Percy G. RuizMamani

111 **Brain imaging provides insights about the interaction between instruction and diagram use for mathematical word problem solving**
Hiroaki Ayabe, Emmanuel Manalo, Noriko Hanaki, Hiroyo Fujita and Michio Nomura

130 **Sample size determination for Bayesian ANOVAs with informative hypotheses**
Qianrao Fu, Mirjam Moerbeek and Herbert Hoijtink

141 **Student conceptual level scale: Development and initial validation**
Xi Yang, Jia Dong and DingLiang Tan

154 **Beyond quality and quantity: Representing empirical structures by embedded typologies**
Matthias Borgstede and Caroline Rau

165 **Higher education management in western regions by educational power strategy and positive psychology**
Xiaomeng Sun

TYPE Editorial
PUBLISHED 13 June 2023
DOI 10.3389/fpsyg.2023.1225744

OPEN ACCESS

EDITED AND REVIEWED BY
Douglas F. Kauffman,
Medical University of the Americas—Nevis, United States

*CORRESPONDENCE
María Soledad Segretin
✉ soledadsegretin@gmail.com

RECEIVED 19 May 2023
ACCEPTED 31 May 2023
PUBLISHED 13 June 2023

CITATION
Segretin MS, Chiu M-S, Tang K-Y, Gutierrez de Blume AP and Declercq C (2023) Editorial: Methods and applications in educational psychology. *Front. Psychol.* 14:1225744. doi: 10.3389/fpsyg.2023.1225744

Editorial: Methods and applications in educational psychology

María Soledad Segretin[1]*, Mei-Shiu Chiu[2], Kai-Yu Tang[3], Antonio P. Gutierrez de Blume[4] and Christelle Declercq[5]

[1]Unit of Applied Neurobiology (UNA, CEMIC-CONICET), National Scientific and Technical Research Concul (CONICET), Buenos Aires, Argentina, [2]College of Education, National Chengchi University, Taipei City, Taiwan, [3]Graduate Institute of Library & Information Science, National Chung Hsing University, Taichung, Taiwan, [4]Department of Curriculum, Foundations, and Reading, Georgia Southern University, Statesboro, GA, United States, [5]Université de Reims Champagne-Ardenne, France, Laboratoire C2S (Cognition, Santé, Société) EA6291, Reims, France

KEYWORDS

educational assessment and evaluation, natural settings, mixed-methods, learning strategies, individual differences, academic performance and behavior, educational resources

Editorial on the Research Topic
Methods and applications in educational psychology

This Research Topic aims to highlight the latest research methods used to investigate fundamental questions in Educational Psychology. The 14 contributions included in this Research Topic bring major conceptual and methodological advances in six Research Topics, in terms of methodology use.

A challenge for this field of research, and the first theme for this Research Topic of articles, addressed methodological issues. These studies use diverse novel methodologies. Magliano et al. examined how theoretically motivated and computationally derived indices of cohesion can be used to explore aspects of coherence-building and how these measures relate to college students' foundational reading skills. Based on the analysis of a large sample, results suggested that the computational analysis of constructed response associated with coherence-building strategies is correlated with individual differences in foundational reading skills. In addition, the relations between foundational reading skill and response cohesion were relatively stable across contexts (i.e., text genre and institutions). Borgstede and Rau propose a new approach to the problem of integration in mixed methods research that builds on a representational understanding of empirical science. From this perspective, qualitative and quantitative modeling strategies constitute two different ways to represent empirical structures. Whereas, qualitative representations focus on the construction of types from cases, quantitative representations focus on the construction of dimensions from variables. Authors argue that types and dimensions should be integrated within a joint representation of the data that equally acknowledges qualitative and quantitative aspects. Fu et al. introduce the R package SSDbain, which can be used to calculate the sample size required to evaluate hypotheses using the Approximate Adjusted Fractional Bayes Factor (AAFBF) for one-way ANOVA models, as implemented in the R package bain. The Bayesian ANOVA, Bayesian Welch's ANOVA, and Bayesian robust ANOVA are available. Using the R package SSDbain and/or the tables provided in this paper, researchers in the social and behavioral sciences can easily plan the sample size if they intend to use a Bayesian ANOVA. Including a neural level of analysis, Ayabe et al. study the development of diagram use competence following the provision of task-appropriate instruction. They focused on both behavioral and neurophysiological evidence (i.e., brain activity, using functional near-infrared spectroscopy or fNIRS). The study was implemented in a sample of children and adolescents (10–19 years of age) and participants were asked to solve mathematical word problems for which the use of tables (which is one kind of diagram) was deemed effective. Results demonstrate important neurophysiological changes resulting from task-appropriate instruction that promotes effective strategy use and improves learning performance.

The second group of articles addressed the role of different individual characteristics in student academic life and learning, using survey and related quantitative data analysis methods. In particular, Shi and Qu analyzed the moderating role of self-monitoring in the association between cognitive abilities (i.e., memory, representational, information processing, logical reasoning, and thinking conversion), and academic achievement in a sample of adolescent students. The results showed that cognitive ability can positively affect academic achievement, while self-monitoring moderate the effect of cognitive ability on academic performance, in particular on mathematics and English subjects. By his part, Zhang et al. implement a survey in a large sample of college students to examine the mediating role of self-concept and social support in the associations between physical exercise and depression. The results suggest that physical exercise negatively predicted college depression. Moreover, self-concept and social support mediate the relationship between physical exercise and depression.

The third group of articles aimed to propose adaptation, validation and/or development of instruments for specific educational assessment. Yang et al. describe the initial development and validation of the Student Conceptual Level Scale with a large sample of students from secondary schools. The authors constructed a three-factor model (learning awareness level, autonomous input level, and environmental coping level), each with its own independent set of items. This study validated the use of full-scale and subscale scores and examined their relationship with different validity criteria (i.e., autonomous learning, mental effort, and academic scores). This updated measure reflects the value and role of the conceptual level in the learning and individual development of students, and provides a more complete frame of reference for the use of the conceptual level in teaching and learning. Gonzales-Valdivia et al. adapted to Spanish and evaluate the psychometric properties of the H-Sat Scale to assess satisfaction with school in Peruvian students between 11 and 18 years old. The scale presented adequate internal consistency for each of the five factors (confirmed through a confirmatory analysis). This measurement tool could be used for the evaluation of interventions in school and health contexts to assess other aspects of wellbeing necessary for their development in school-age students.

The fourth group of articles addressed conceptual or theoretical models or frameworks in the field. Jirout et al. highlighted the importance of curiosity in children's development and academic performance, which has been widely acknowledged but not extensively studied in classroom settings. The result of the feasibility test showed that the Curiosity in Classrooms (CiC) framework is a useful tool that can help teachers to support students' development and academic success. The authors begin by describing the framework they used to identify specific instructional practices that can promote curiosity, and then focus on the development of a coding protocol for and test of the framework. Sun proposed the construction of an educational resource recommendation model based on Takagi and Sugeno (T–S) fuzzy neural network, verified the feasibility of the model, combined the educational resource recommendation model with university teaching, and analyzed the application effect. After applying the educational resource recommendation model, the accuracy of educational resource recommendation is improved, and the design is feasible. The educational management mode with positive psychological emotions has a good teaching effect, which can improve teachers' dedication and concentration. Teaching resource recommendation mode can improve college students' interest in the application of teaching resources to a certain extent, and their application satisfaction is improved. Based on learning strategy models and using thematic analysis, Liu and Uesaka proposed a new lecture note-taking framework comprising shallow and deep lecture note-taking. For that, data from high school students from two countries (Japan and China) were analyzed to explore cognitive activities in which students engage when taking lecture notes in mathematics classes. Results provided insights into the cognitive activities accompanying lecture note-taking, such as the metacognitive function, which has yet to be explored in previous research. In addition, the authors inferred that the differences in lecture notes might result from the influence of the student's beliefs and teachers' instruction styles.

The fifth, Annamalai et al. used a qualitative case study approach to understand the factors that affect students' motivation and satisfaction with online assessments in three different countries (Malaysia, Lithuania, and Spain) during the COVID-19 pandemic. The finding highlighted the importance of effective assessment guidelines and feedback in online learning environments. In addition, the impact of technology is also suggested for the investigation in the contexts of online assessment, such as virtual reality or gamification, and how they impact student motivation and engagement.

The sixth, two literature review methodologies are used to review papers in this Research Topic. Shaojie et al. discuss, based on a systematic literature review, the effects of audiovisual input on second language acquisition and the factors that influence the difficulty of audiovisual learning. The authors concluded that audiovisual input could provide more authentic language input and more adequate and richer multimodal cultural and situational contexts, which can better promote learners' understanding of the content and stimulate learners' interest in participating in listening comprehension tasks. Factors affecting audiovisual multimodal input difficulty included subtitles, video inputs, and sounds and pictures relationships. Lin et al. review the research trends of scaffolding in the field of science education. Then, descriptive and co-word analysis were conducted to examine the selected articles published in the Social Science Citation Index journals in the past 20 years. Overall, this study reveals a growing trend of science educators' academic publications about scaffolding in the recent two decades. In addition, results showed that "scaffolding," "support," and "design" were the top three most frequently used keywords during 2000 and 2019. Visualization of co-word networks in each 5-year period further helps clarify both educators' common research focuses and relevant research trends.

In summary, the 14 papers included in this Research Topic focus on six major Research Topics and using diverse methodologies. The first group has four papers, which directly deal with the issue of methodologies in the field, including natural language processing, mixed-methods integration, sample size determination, and behavioral -neurophysiological data integration. The second group (with two papers) focuses on students' academic life, using survey as the major methodology. The two papers in the third group focuses on scale development. The three papers in the fourth group focuses on framework building and further exploration or examination. One paper uses a qualitative, interview method, and two papers using two different literature review methods. This Research Topic collection may mimic the present research fronts and suggest developing more diverse, novel methodologies to study topics in relation to learning, instruction, and assessment in educational psychology. Multi-, inter-, and transdisciplinary endeavor may be needed to advance a deeper understanding of the field and to face human challenges in the world (e.g., artificial intelligence and the pandemic).

Author contributions

MSS and M-SC: writing original draft. K-YT, APGB, and CD: writing review. All authors contributed to the article 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

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

REVIEW
published: 18 February 2022
doi: 10.3389/fpsyg.2022.844425

A Co-word Analysis of Selected Science Education Literature: Identifying Research Trends of Scaffolding in Two Decades (2000–2019)

Tzu-Chiang Lin[1,2], Kai-Yu Tang[3], Shu-Sheng Lin[4], Miao-Li Changlai[5] and Ying-Shao Hsu[6,7]*

[1] Center for Liberal Arts, National Kaohsiung University of Science and Technology, Kaohsiung City, Taiwan, [2] Center for Teacher Education, National Kaohsiung University of Science and Technology, Kaohsiung City, Taiwan, [3] Department of International Business, Ming Chuan University, Taipei City, Taiwan, [4] Graduate Institute of Mathematics and Science Education, National Chiayi University, Chiayi City, Taiwan, [5] Department of Digital Multimedia Design, China University of Technology, Taipei City, Taiwan, [6] Graduate Institute of Science Education, National Taiwan Normal University, Taipei City, Taiwan, [7] Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taipei City, Taiwan

OPEN ACCESS

Edited by:
Antonio P. Gutierrez de Blume, Georgia Southern University, United States

Reviewed by:
Mario Mäeots, University of Tartu, Estonia
Gwo-Jen Hwang, National Taiwan University of Science and Technology, Taiwan

****Correspondence:***
Ying-Shao Hsu
yshsu@ntnu.edu.tw

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

Received: *28 December 2021*
Accepted: *20 January 2022*
Published: *18 February 2022*

Citation:
Lin T-C, Tang K-Y, Lin S-S, Changlai M-L and Hsu Y-S (2022) A Co-word Analysis of Selected Science Education Literature: Identifying Research Trends of Scaffolding in Two Decades (2000–2019). Front. Psychol. 13:844425. doi: 10.3389/fpsyg.2022.844425

This study aims to identify research trends of scaffolding in the field of science education. To this end, both descriptive analysis and co-word analysis were conducted to examine the selected articles published in the Social Science Citation Index journals from 2000 to 2019. A total of 637 papers were retrieved as research samples through rounds of searching in Web of Science database. Overall, this study reveals a growing trend of science educators' academic publications about scaffolding in the recent two decades. In these sample papers, from 1,487 non-repeated keywords, we extracted 286 author-defined keywords shared by at least two studies as a benchmark dictionary. A series of co-word analyses were then conducted based on the dictionary to reveal the underlying co-occurring relationships of the words in title and abstract of the sample papers. Results showed that "scaffolding," "support," and "design" were the top three most frequently used keywords during 2000 and 2019. Visualization of co-word networks in each 5-year period further helps clarify both educators' common research foci and relevant research trends. Derived discussion and potential research directions are also provided.

Keywords: co-word analysis, research trends, scaffolding, science education, educational technology

INTRODUCTION

In the last decades, educators have reached a consensus that scaffolding analogises learners' knowledge building with proper supports. Wood et al. (1976) first identified young children's development progression with learning scaffolds provided by their parents. To accomplish tasks in daily life, young children should be equipped with certain skills and corresponding confidence in themselves. In this progression, learning supports act as scaffolding that guides children's advancement beyond what they may achieve simply on their own. With decreasing aids, usually referred to as fading, children may gradually perform tasks in a more self-directed manner. This provides insight into educational reforms from the perspective of instruction practices in the classroom (Bliss et al., 1996). To facilitate meaningful learning, scaffolding is deemed as learner-centered guidance that affords goal setting, knowledge construction, as well as self-reflection during the learning process (Davis and Linn, 2000; Azevedo et al., 2004).

Scaffolding hence reflects ideas of social constructivism (Vygotsky, 1978; Bliss, 1994) and the zone of proximal development (ZPD) (Rogoff, 1990; Metz, 1995, 1997). The set of existing theories depicts instructional strategies that may overcome the gap between what students learn on their own and what they learn with assistance. In the field of science education, the theories have also been shown to fundamentally align with relevant learning approaches such as argumentation (McNeill et al., 2006; Belland et al., 2011; Noroozi et al., 2018), project-based learning (Land and Zembal-Saul, 2003; Reiser, 2004), problem-based learning (Belland et al., 2011; Kim and Hannafin, 2011), as well as inquiry (Reiser, 2004; Hsu et al., 2015). Science educators have exerted impressive efforts to apply scaffolding in sophisticated learning contexts that highlight students' metacognition, higher order thinking skills, and modeling practice (Fretz et al., 2002; Tang et al., 2016; Toleda and Dubas, 2016; Alrawili et al., 2020). Moreover, the development of educational technology has enabled researchers and teachers to accomplish more expected scaffolds in instructional designs (Quintana et al., 2004; Oh and Jonassen, 2007; Kim and Hannafin, 2011). In sum, in the field of science education, scaffolding is an issue that continues to attract educators' attention. Understanding the relationships interwoven with such issue through systematic reviews of the relevant literature may inform educators of the trends and research spaces.

Lin T. C. et al. (2012) demonstrated a review work on studies that focused on the science education field. Their review examined empirical studies published in Social Science Citation Index (SSCI) journals from 1995 to 2009 to identify research trends regarding scaffolding. Most of the studies focused on improving contexts for science learning. However, an unexpected finding was that in the 15 years, researchers invested only limited efforts in unveiling how science teacher education improved teachers' professional development related to scaffolding. It was also surprising that fading, an essential part of the scaffolding process, was precisely described in <10% of the studies. Researchers thereafter reported similar review works regarding scaffolding that addressed various research domains such as literacy learning (Brownfield and Wilkinson, 2018), computer-based scaffolding in STEM education (Belland et al., 2017) as well as metacognitive scaffolding for online information searching (Zhou and Lam, 2019). With regard to technology-enhanced science education, Wu et al. (2021) reviewed 60 studies to identify research trend of technology-enhanced chemistry learning. Chen et al. (in press) targeted 44 academic publications to analyze and report implications of flipped science learning. The previous review works with content analysis on one hand revealed possible research directions, but on the other hand highlighted potential difficulties in certain research issues. Consecutive trends and alternative analytic techniques may be necessary to expand a more systematic view of development in the research field.

Comparing with content analysis used in previous review studies (Lin T. C. et al., 2012; Brownfield and Wilkinson, 2018; Zhou and Lam, 2019), a co-word network analysis potentially provides complementary views to explore the intellectual structure of keywords used by researchers. It is also noteworthy that the process of manual reviews in the content analysis are time-consuming. Researchers may encounter difficulties when they try to examine the research trends based on an overly large quantity of literature or literature from an overly long period of time. Literature review approach such as content analysis on large-size texts in nature requires imaginable efforts. It is almost impossible to consider information that possibly exists beyond coding framework pre-established by researchers. The bibliometric method may be a satisfactory approach that can unveil complex sets of relationships among a large body of literature. Bibliometric analysis stems from library science, and basically taxonomizes literature with certain characteristics. Such research approach is especially important today as the amount of academic literature has mushroomed in recent years with unimaginable speed. Co-word analysis, a kind of bibliometric method, was hence developed to calculate the co-occurrence counts of selected words in literature (Callon et al., 1991). Recently, researchers in education have adopted the method to explore the research themes in educational research (Huang et al., 2020), STEAM literature (Marín-Marín et al., 2021), and intellectual structure of a SSCI-journal's publications (López-Belmonte et al., 2021). In this study, we adopted a series of co-word network analyses to provide visualized network structure of the scaffolding keywords based on co-occurrence relationships.

Furthermore, using author-defined keywords to represent the focal interest of a selected article is valid in the current analysis as the main idea of validity refers to how accurately a method measures what it is intended to measure (Kelley, 1927). Researchers have suggested that content validity is a key for the valid measure (Rubio et al., 2003). Therefore, this study conducted a co-word network analysis of literature to analyze the structure of keyword co-occurrence to uncover the development trends of empirical studies about scaffolding in science education. We especially focus on the thorough understanding of mutual wording that educators inclined to identify their research focuses, approaches, and findings about scaffolding. The keywords designated by authors are especially deemed as key clues to understand how the research content was defined (Assefa and Rorissa, 2013). A co-word analysis of the status and degree of keyword sharing may represent a precise picture of core research foci in the literature.

The review in this study aimed to identify relationships of high-quality academic articles published in Social Science Citation Index (SSCI) journals within the recent 20 years. This study also aimed at revealing ideas associated with scaffolding. Further interpretations of similarities and dissimilarities that science educators regarded in the contemporary research on scaffolding are, therefore, highlighted. The specific research questions are as follows:

1. What keywords regarding scaffolding were designated in the 2000–2019 SSCI journal articles? What is the trend of these keywords for each 5-year period?

2. What is the visualized pattern of co-word analysis of the research articles regarding scaffolding published in the 2000–2019 SSCI journals? What is the trend of the patterns for each 5-year period?

DATA AND METHODS

Data

This study conducted several rounds of topic searches on the Web of Science (WoS) database using the keywords "scaffold" and "scaffolding" coupled with other keywords including science, physics, chemistry, biology, earth science, mathematics, environmental science, medical science, STEM, and STEAM. A total of 1,106 papers published in SSCI journals from 2000 to 2019 were retrieved. A series of relevance checks was conducted to filter out some irrelevant papers for the subsequent analysis. First of all, redundant papers from different queries were excluded ($n = 401$). Next, in line with previous review studies (Lin T. C. et al., 2012), we removed some editorial materials ($n = 5$) and correction data ($n = 1$), and only retained article- and review-type papers in the dataset. Note that some incomplete data and some non-English papers ($n = 4$) were also removed. Last, the keywords used for the topic search were again used to examine the relevance of all remaining papers. Those papers only indexed by the system-defined KeywordPlus (a special category defined by the WoS) were removed ($n = 57$). As a result, a total of 637 papers remained for the subsequent analysis. The sample papers cited in this article are asterisked in the reference list.

The Identification of Keywords

There are many different standards for the presentation of keywords among various journal publications. For example, some journals only allow authors to select keywords from build-in datasets instead of using author-defined keywords, while some other journals publish papers without keywords. In this present study, we found that the keyword column of over 24.5% (156 out of 637 papers) of the articles was empty. To solve this problem, we present the three-step procedure we adopted to re-index all 637 papers.

Dictionarizing Co-words

First, we dictionarized the available author-defined keywords as the reference to re-index words presented in title and abstract of the selected articles. All the author-defined keywords from 481 papers were collected. As a result, we found a total of 1,487 non-repeated keywords. This set of keywords represents the focus knowledge of the scaffolding research and was treated as the boundary of the keywords used in the field. Second, the frequency of use of every keyword was counted. Among all 2,365 usage counts, approximately one half ($n = 1{,}166$, 49.3%) were focused on a total of 286 author-defined keywords that were used by at least two researchers in the field. Third, this set of 286 keywords was deemed as a benchmark dictionary to re-index all of the analyzed papers. This process was performed to construct a standardized set of keywords to characterize all 637 papers. In line with the practice of topic searches in the WoS, we selected the two columns of article title and abstract for the process of dictionarization. This also avoided the potential interference of simply dictionarizing the whole article. A co-word analysis was used to detect whether a paper incorporated any of the abovementioned 286 keywords. As a result, all 637 articles were re-indexed with a range of 2–31 dictionarized keywords. Analysis of the keyword structure which evolved in each period of interest from 2000 to 2019 is further discussed based on the perspective of co-word network analysis.

Co-word Network Analysis

A co-word analysis assumes that the author-defined keywords in the academic publications are the key description of its research content (Dehdarirad et al., 2014; Huang et al., 2020). In this manner, the links between a large number of co-occurring keywords represent the focal interest of the Research Topic in the field. By definition, two keywords of interest that co-occur within the same paper suggest a certain degree of bibliometric relationship between the topics to which the keywords refer (Cambrosio et al., 1993). For example, keywords A and B co-occur if they are used together in a focal research paper. According to Leydesdorff (1989), however, "a word which occurs only once cannot form a co-word linkage" as it may appear occasionally or be determined by researchers in a single study. In the current review, only the co-occurrence links that outperformed the average strength were considered as influential keywords of scaffolding research.

To calculate the co-occurring relationship among keywords, a co-occurrence was formed. In this study, a total of 286 keywords were listed in both rows and columns, where each cell represents the frequency of co-occurrence counts of the two focal keywords. The more frequently the keywords were co-used by the researchers, the higher likelihood that those two keywords represented similar concepts in the context of scaffolding research. On the other hand, two keywords were considered as dissimilar from each other if they were not used together by field researchers. To further observe the co-occurrence network in the scaffolding research, some networking measures of centrality (e.g., degree centrality measures) were adopted to identify the most influential keyword corpus in the field. After the co-occurrence matrix was obtained, VOSviewer version 1.6.15 was used to visualize the co-occurrence structure of the keyword corpus in each period of research from 2000 to 2019.

RESULTS AND DISCUSSION

Descriptive characteristics and research trends regarding scaffolding of the articles published in SSCI journals were revealed through a series of co-occurring keyword analyses and comparisons. Thereafter, we present the results of co-word network analysis to visualize relationships among the keywords in different periods with a 5-year interval from 2000 to 2019.

Descriptive Characteristics of the Articles

Figure 1 presents the number of articles on scaffolding published in each year from 2000 to 2019. The published articles on

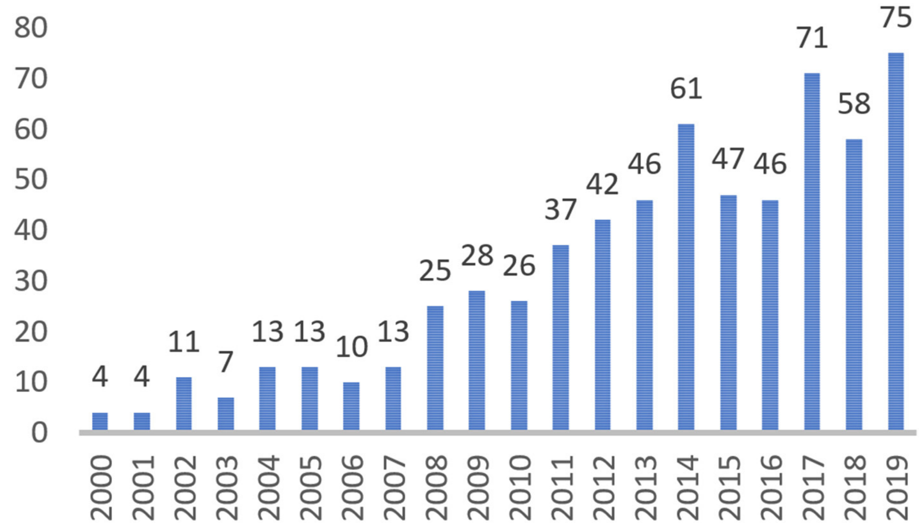


FIGURE 1 | Numbers of the articles on scaffolding from 2000 to 2019.

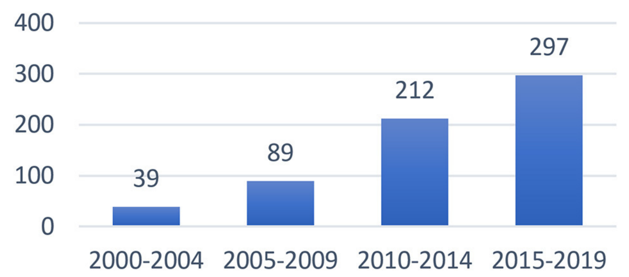


FIGURE 2 | Numbers of the articles on scaffolding in 5-year intervals.

scaffolding continued to increase, which indicated that the field was growing and attracting more researchers to contribute their efforts, especially during 2013–2014, 2016–2017, and 2018–2019. The published articles on scaffolding at 5-year intervals also showed an increasing trend. A closer look at the trend analysis indicated an obvious jump from 2005–2009 to 2010–2014 (**Figure 2**).

Besides, when focusing on science education and educational technology fields, the numbers of selected articles published in the nine journals (**Table 1**) related to science education and in the 14 journals (**Table 2**) related to educational technology increased over the 5-year intervals (**Figures 3**, **4**). The trend analysis indicated an obvious jump from 2005–2009 to 2010–2014.

It should be noted that the jumps occurred in varying sets of trend analyses. The jumps could be because a growing number of researchers valued the importance of scaffolding and explored its effects on learning and teaching after the National Science Council (NRC, USA) released the final Framework for K—12 Science Education in 2011, and the NGSS (Next Generation Science Standards) were published on its website in 2013.

TABLE 1 | The selected science-education journal list.

Rank*	Journal title	Total cites	Impact factor	Eigenfactor score
10	Journal of Research in Science Teaching	6,518	3.870	0.004540
12	Studies in Science Education	718	3.700	0.000590
19	Science Education	5,048	3.500	0.002800
66	Research in Science Education	1,728	2.248	0.001420
100	International Journal of STEM Education	368	1.850	0.000910
133	International Journal of Science and Mathematics Education	1,358	1.578	0.001830
146	International Journal of Science Education	5,613	1.485	0.004160
212	Journal of Baltic Science Education	385	0.915	0.000260
248	Cultural Studies of Science Education	530	0.437	0.000550

**Nine science-education journals selected from top 250 in the "Education and Educational Research" category of 2019 Journal Citation Report in WoS database.*

Trend Analysis of the Keywords and Co-occurring Keywords in the Selected Articles

The results in **Table 3** show the top 30 keyword frequencies (about half of the 63 keywords, the frequencies of which were above the average of 23.5) and the derived trends. The total frequency of "scaffolding" is shown as 611 instead of 637 because there were 26 articles which used scaffolding in compound phrases such as computer-based scaffolding (12), distributed scaffolding (4), and metacognitive scaffolding (10). Referring to the trends in **Table 3**, the highest occurring keywords in

TABLE 2 | The selected educational-technology journal list.

Rank*	Journal title	Total cites	Impact factor	Eigenfactor score
3	Internet and Higher Education	3,217	6.566	0.003060
4	Computers & Education	15,521	5.296	0.013370
9	International Journal of Computer-Supported Collaborative Learning	796	4.028	0.000740
27	International Journal of Educational Technology in Higher Education	371	3.080	0.000450
31	British Journal of Educational Technology	4,359	2.951	0.003410
36	IEEE Transactions on Learning Technologies	948	2.714	0.001160
59	International Review of Research in Open and Distributed Learning	2,443	2.297	0.002030
72	Journal of Educational Computing Research	1,475	2.180	0.001030
79	Journal of Computer Assisted Learning	2,352	2.126	0.002050
83	Educational Technology and Society	3,136	2.086	0.003310
95	Australasian Journal of Educational Technology	1,468	1.956	0.000990
96	Interactive Learning Environments	1,197	1.938	0.001940
124	Journal of Science Education and Technology	1,876	1.644	0.001770
191	Research in Science and Technological Education	501	1.111	0.000450

**14 educational-technology journals selected from top 250 in the "Education and Educational Research" category of 2019 Journal Citation Report in WoS database.*

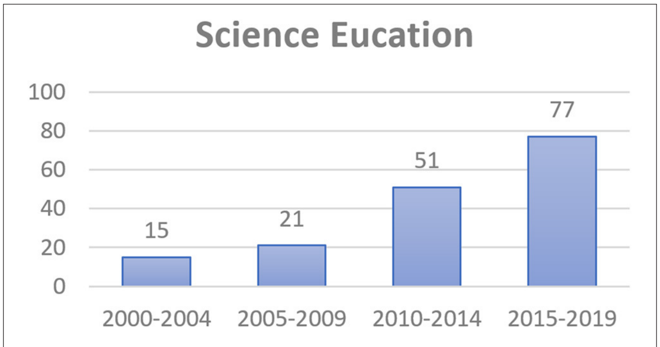


FIGURE 3 | Numbers of the articles published in the science-education journals.

the selected articles continued to increase from the first period (2000–2004) to the fourth period (2015–2019). The numbers also indicated a jump from the second period (2005–2009) to the third period (2010–2014). This revealed the same trend with the total number of scaffolding-related articles (**Figure 2**), the articles published in the journals related to science education (**Figure 3**), and the articles published in the journals related to educational technology (**Figure 4**) due to the release of the Framework for K-12 Science Education in 2011 and NGSS in 2013.

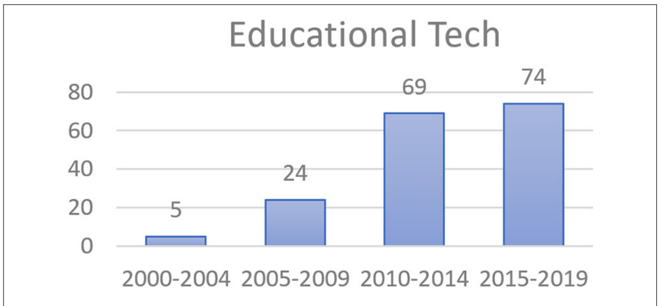


FIGURE 4 | Numbers of the articles published in the educational-technology journals.

For better understanding of the research themes of the selected articles, we classified the keywords into six groups. First are the keywords related to scaffolding characteristics such as support (307), design (256), context (156), learning environment (116), interaction (116), technology (93), computer (67), and feedback (54). Second are the keywords which indicated scaffolding for specific learning performances such as knowledge (192), practice (156), cognition (148), experiment (81), evidence (78), reasoning (71), observation (66), modeling (61), evaluation (59), reflection (55), and argumentation (54). Third are the keywords about teaching (who and when to design and provide scaffolding) such as teachers (147), and in instruction (177) and assessment (82). Fourth are the keywords related to learning disciplines to be learned such as STEM (138), Science (117), mathematics (67), and physics (65). Fifth are the keywords indicating the grades of learners such as high school (138) and elementary school (57), which were two major groups of subjects in the selected studies. Sixth is the most popular methodology used, which was case study (58). A keyword not included in the six groups was science education (76), which is a general keyword used by science educators in their articles.

Even though the trend analysis showed that the three most frequent keywords (scaffolding, support, and design) were the same throughout the entire 20 years, other high-frequency keywords occurred differently in the different 5-year periods. When we selected 33 as the threshold number of high-occurrence keywords, which was one S.D. (9.7) from the frequency average (23.5) of the occurrence frequency, no keywords other than scaffolding occurred (37) in the first period (2000–2004). Regarding keywords that appeared more than 33 times in the second period (2005–2009), only the three previously mentioned most frequent keywords related to scaffolding characteristics were identified, namely scaffolding (87), support (38), and design (34). Regarding an occurrence frequency of more than 33 times in the third period (2010–2014), six keywords related to scaffolding characteristics occurred: scaffolding (207), support (106), design (87), context (57), interaction (34) and learning environment (39); three keywords related to teaching were instruction (56), teacher (47), and assessment (37); three keywords related to learning performance were knowledge (60), cognition (49), and

TABLE 3 | The most frequently used keywords in scaffolding research.

Rank	Keywords	2000–2004	2005–2009	2010–2014	2015–2019	Total
1	Scaffolding	**37**	**87**	**207**	**280**	611*
2	Support	17	**38**	**106**	**146**	307
3	Design	16	**34**	**87**	**119**	256
4	Knowledge	11	**32**	**60**	**89**	192
5	Instruction	10	31	**56**	**80**	177
6	Context	12	22	**57**	**65**	156
7	Practice	8	25	**46**	**77**	156
8	Cognition	7	23	**49**	**69**	148
9	Teacher	16	14	**47**	**70**	147
10	STEM	7	12	**55**	**67**	141
11	High school	4	20	**41**	**73**	138
12	Science	9	28	**34**	**46**	117
13	Interaction	13	16	**34**	**53**	116
13	Learning environment	7	21	**39**	**49**	116
15	Technology	6	12	29	**46**	93
16	Assessment	4	8	37	**33**	82
17	Experiment	3	9	24	**45**	81
18	Evidence	3	8	25	**42**	78
19	Science education	4	5	31	**36**	76
20	Reasoning	2	13	20	**36**	71
21	Computer	4	7	21	**35**	67
21	Mathematics	4	6	27	30	67
23	Observation	3	11	25	27	66
24	Physics	1	8	26	30	65
25	Modeling	4	10	14	**33**	61
26	Evaluation	6	6	13	**34**	59
27	Case Study	4	6	16	32	58
28	Elementary school	7	7	22	21	57
29	Reflection	3	8	12	32	55
30	Argumentation	5	8	15	26	54
30	Feedback	2	7	23	22	54

**12 articles in computer-based scaffolding, 4 articles in distributed scaffolding, and 10 articles in metacognitive scaffolding (total = 637).*
The bold font indicates the occurrence frequency of more than 33 times (one S.D. from the frequency average).

practice (46); and two keywords related to disciplines were STEM (55) and science (34). Regarding an occurrence frequency of more than 33 times in the fourth period (2015–2019), one more keyword (technology) related to scaffolding characteristics occurred in addition to the six keywords in the third period; the same three keywords related to teaching as the third period appeared; five more keywords (experiment, evidence, reasoning, modeling, and evaluation) related to learning performances occurred in addition to the same three high-occurrence keywords as the third period; the same two keywords (STEM and science) related to disciplines as in the third period occurred. Compared to the first three periods, more interest in technology was found together with scaffolding in the fourth period. Moreover, inquiry practices (experiment, evidence, reasoning, and modeling) and higher order thinking skills (reasoning, modeling, and evaluation) seemed to attract attention in the most recent period (2015–2019).

TABLE 4 | The frequently co-occurred keywords in the selected science education journals.

Rank	Keywords	2000–2004	2005–2009	2010–2014	2015–2019	Total
1	Scaffolding	15	20	**51**	**73**	159
2	Support	6	9	18	**39**	72
3	Science education	6	0	**26**	**32**	64
4	Knowledge	6	10	19	**24**	59
5	Design	5	8	18	**26**	57
6	Teacher	7	8	14	**25**	54
7	Science	5	9	16	19	49
8	Practice	1	7	12	**24**	44
9	Context	6	3	14	17	40
10	STEM	1	3	12	23	39
11	Instruction	3	4	12	17	36
12	Evidence	3	2	13	17	35
13	Cognition	4	2	9	14	29
14	Reasoning	0	3	8	14	25

The bold font indicates the co-occurred times of more than the frequency average 23.5.

Comparisons of Co-occurring Keywords in Science-Education Journals and Educational-Technology Journals

Table 4 shows the top 14 frequently co-occurring keywords with co-occurrence times above 23.5 as the average frequency in the selected science education journals. The frequencies of the top co-occurring keywords continued to increase from the first period (2000–2004) to the fourth period (2015–2019). It is worth noting that the trend showed a significant increase from the third period (2010–2014) to the fourth period (2015–2019). Scaffolding hence acted as the most frequently co-occurring keyword during the four periods. When we used 24 (the frequency average was 23.5) as the threshold number of frequently co-occurring keywords, no co-occurring keywords were included in the first period (2000–2004) or in the second period (2005–2009). Regarding 24 or more times of co-occurring frequency in the third period (2010–2014), only two co-occurring keywords were identified, namely scaffolding (51) and science education (26). Compared with the same two co-occurring keywords in the third period, five more co-occurring keywords were identified: support (39) and design (26) related to scaffolding characteristics, teacher (25) related to teaching, and knowledge (24) and practice (19) in the fourth period (2015–2019).

The top 18 frequently co-occurring keywords of the articles published in the selected educational technology journals are presented in **Table 5**. The co-occurrence times of the keywords are more than the average number 23.5. The trend indicated a jump from the second period (2005–2009) to the third period (2010–2014). Scaffolding, support, and design were the top three frequently co-occurring keywords from the second period (2005–2009) to the fourth period (2015–2019). It was not until the

TABLE 5 | The frequently co-occurred keywords in the selected educational technology journals.

Rank	Keywords	2000–2004	2005–2009	2010–2014	2015–2019	Total
1	Scaffolding	5	**24**	**66**	**67**	162
2	Support	1	12	**40**	**40**	93
3	Design	3	10	**33**	**44**	90
4	Knowledge	1	8	23	**29**	61
5	Cognition	2	10	**26**	20	58
6	Science education	0	8	20	**28**	56
7	Technology	1	8	21	**25**	55
8	Science	3	6	21	21	51
9	STEM	3	4	**24**	20	51
10	Learning environment	3	8	20	16	47
11	Teacher	1	1	21	20	43
12	Context	1	8	15	17	41
13	Interaction	1	4	17	17	39
14	Instruction	2	8	15	13	38
15	Practice	1	5	13	13	32
16	Computer	1	4	13	13	31
17	Assessment	1	3	11	15	30
18	Experiment	0	6	5	15	26

The bold font indicates the co-occurred times of more than the frequency average 23.5.

second period (2005–2009) that the frequency of co-occurring keywords achieved 24. In addition to scaffolding (66), support (40), and design (33) related to scaffolding characteristics, cognition (26) related to learning performances, and STEM (24) related to learning disciplines were included in the third period (2010–2014). Compared with the same top three co-occurring keywords in the third period, three more co-occurring keywords were included: knowledge (29) related to learning performances, science education (28), and technology (25) related to scaffolding characteristics in the fourth period (2015–2019).

According to **Tables 4**, **5**, six co-occurring keywords (technology, interaction, learning environment, computer, assessment, and experiment) were used in article published in science education journals. Two co-occurring keywords (evidence and reasoning) were absent from articles published in educational technology journals throughout the entire 20-year period.

Results of Co-word Network Analysis

To further uncover the trends of research themes in scaffolding research from 2000 to 2019, we conducted a series of CNA to depict the literature in the recent 20 years. **Figure 5** shows the overview of the co-word analysis within the period of 2000–2019. It should be noted that the size of the node represents the frequency count of the keywords, while the thickness of the line indicates the strength of the co-occurrence of the keywords. The figure visualizes the aforementioned research themes regarding scaffolding that the researchers were interested in during the past decades.

With regard to the threshold number of frequently co-occurring keywords as 23.5, the co-occurrence frequencies found in the period of 2000–2004 were insufficient to conduct CNA. This implies that the mutual research interests regarding scaffolding in this period were still divergent. This may have hence led the articles to share relatively fewer co-words.

As for the period 2005–2009 (**Figure 6A**), seven keywords emerged in the co-occurrence network of the scaffolding research. Among these seven words, the links between "scaffolding and science" (the co-word frequency is 51), "scaffolding and design" (42), and "scaffolding and support" (41) are relatively higher than those between "scaffolding and knowledge" (36), "scaffolding and instruction" (32), and "scaffolding and practice" (27). Obviously, how to support students' science learning through designing proper scaffolding is one of the important issues in the relevant research field.

There were 22 keywords related to scaffolding which emerged from the co-word analysis of scaffolding research during 2010–2014, relatively much more than in 2005–2009 (**Figure 6B**). In addition to the higher frequency of the four co-words mentioned in 2005–2009 (scaffolding, science, support, and design), stronger links of the other co-words existed between the words "scaffolding and teacher" (93) and "scaffolding and instruction" (63), which belong to the research theme teaching; and "scaffolding and knowledge" (68), "scaffolding and practice" (53), and "scaffolding and cognition" (51), which belong to the research theme learning performances. Other stronger links of co-words also existed between "scaffolding and STEM" (65) and "scaffolding and context" (57). It is worth noting that the word "teacher" highly co-occurred with the word "scaffolding" compared to the other words. This might mean that the follow-up studies paid much more attention to teachers' roles in the design or use of scaffolding.

Compared with **Figures 6A–C** shows that more keywords ($N = 29$) and more complex links were identified in the co-word analysis of scaffolding research during 2015–2019. The four co-words (i.e., scaffolding, science, support, and design) still rank as the highest. The relationships among other highly co-occurring keywords mentioned in the period 2010–2014 were also enhanced. Especially, the co-words belonging to the research theme scaffolding characteristics, such as technology, computer, interaction, learning environment, and experiment increased greatly during this period. Some of the co-words belonging to the research theme learning performances, such as reasoning, evidence, argumentation, and evaluation, obviously appeared. It was revealed that in this period researchers seemed to focus more on the design of scaffolding enriched by technology or computers to improve students' higher order thinking skills. This also implies that researchers commonly appreciated the necessity to overcome teachers' difficulties to provide proper scaffoldings for supporting students' higher order thinking skills.

Co-word Network Analysis Related to Participants

In the selected articles published in the recent two decades, there existed co-occurrence relationships among scaffolding,

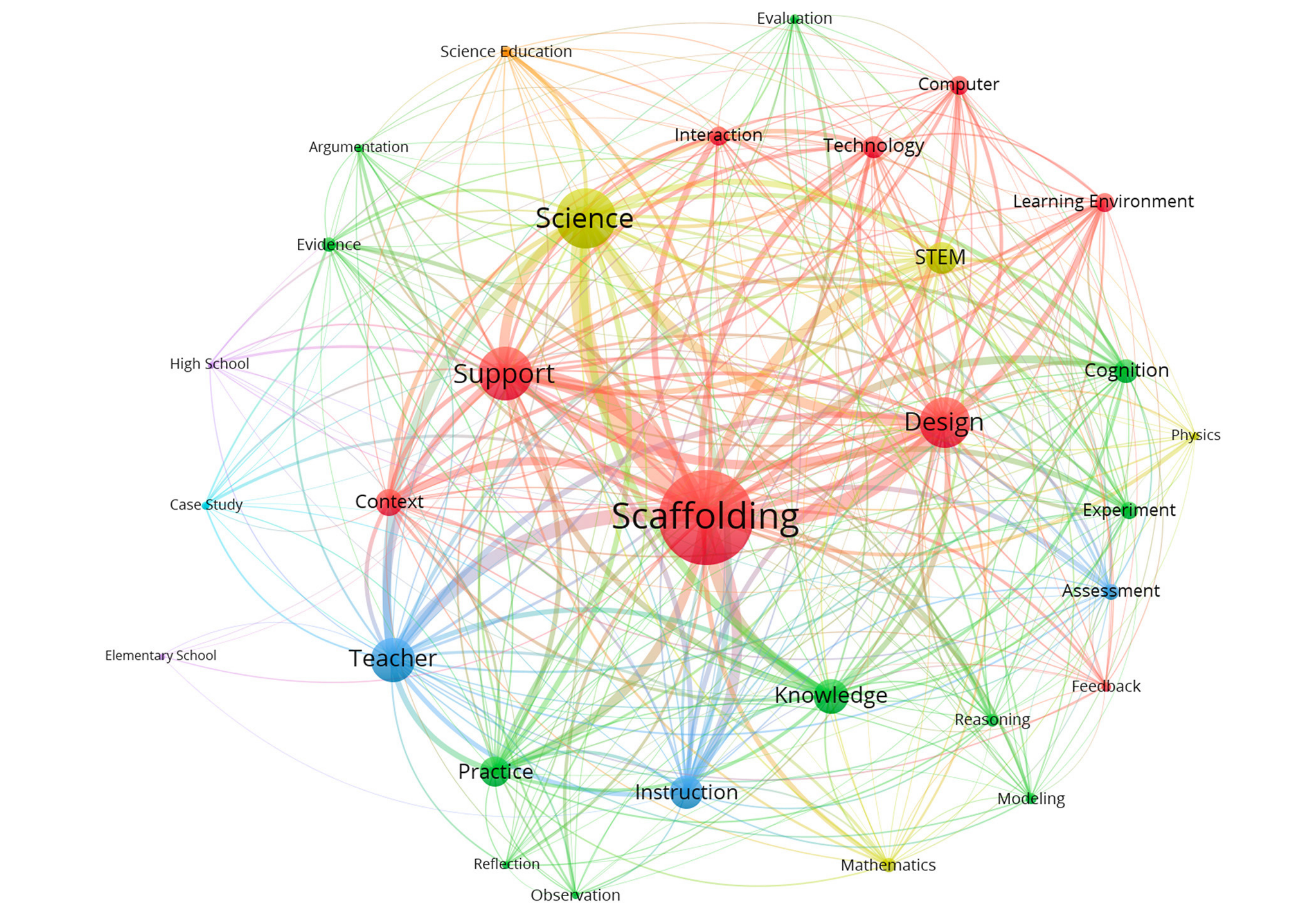


FIGURE 5 | The co-occurred network of scaffolding research (All period: 2000–2019, keywords = 31).

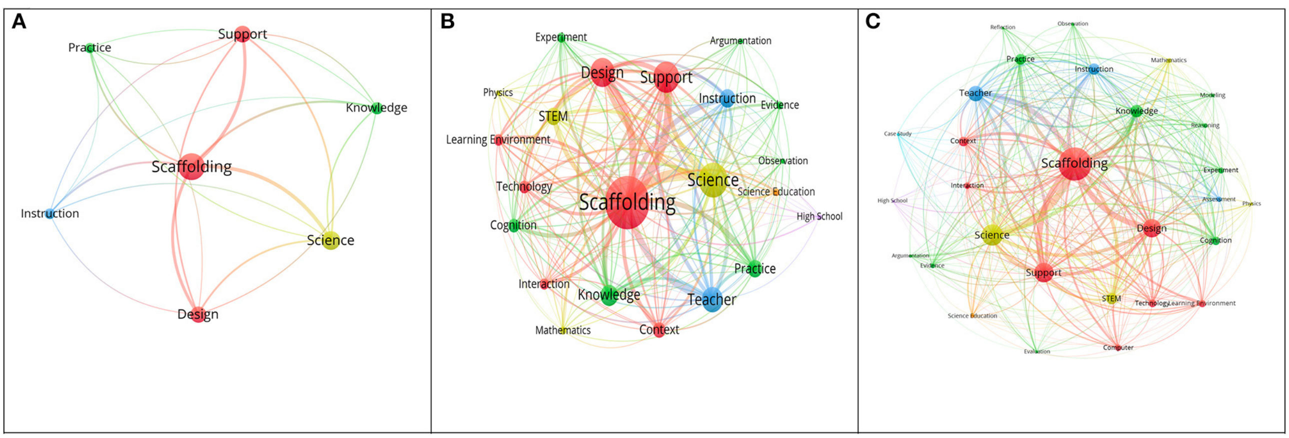


FIGURE 6 | (A) The co-occurred network of scaffolding research (2005–2009, keywords = 7). **(B)** The co-occurred network of scaffolding research (2010–2014, keywords = 22). **(C)** The co-occurred network of scaffolding research (2015–2019, keywords = 29).

support, teacher, instruction, science, and elementary school (**Figure 7A**). Researchers who aimed at scaffolding elementary school learners may specifically tend to address the role of scaffolding characteristics, teaching, and the science discipline itself in their studies (Baker et al., 2008; Kershner et al., 2010; Choi et al., 2015; Decristan et al., 2015; van Uum et al.,

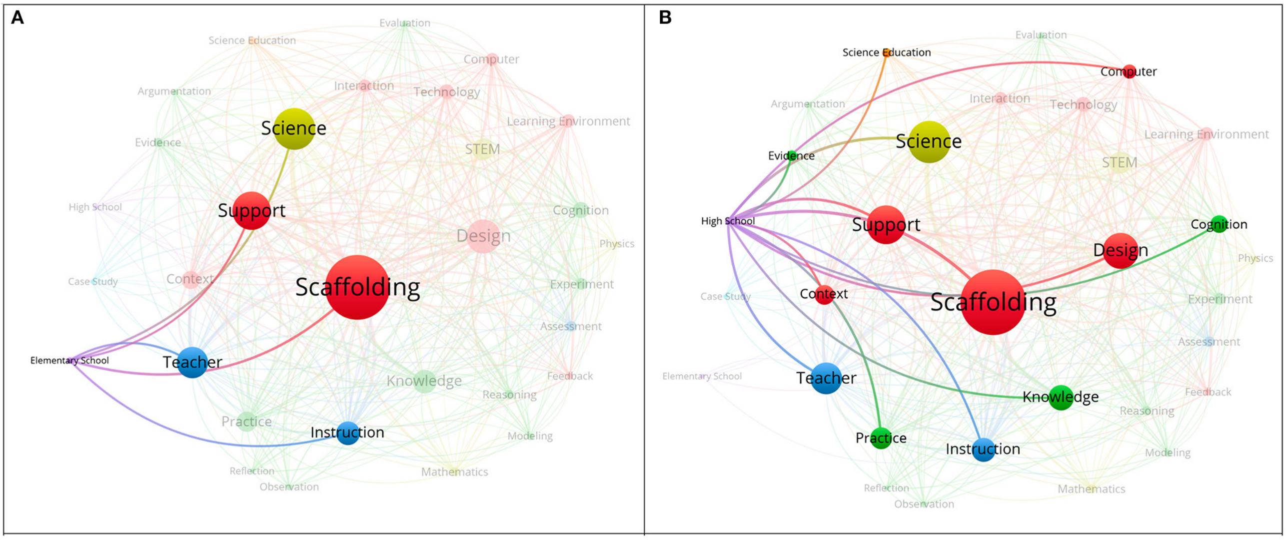


FIGURE 7 | (A) "Elementary school" and other frequently co-word network analysis (keywords = 5). **(B)** "High school" and other frequently co-word network analysis (keywords = 13).

2017). By contrast, more complicated co-word relationships are found in the network associated with scaffolding in high school context (**Figure 7B**). More keywords include scaffolding, support, design, context, and computer related to the research theme of scaffolding characteristics; teacher and instruction related to the research theme of teaching; knowledge, practice, cognition, and evidence related to the research theme of learning performances; science related to the research theme of learning disciplines; as well as science education. It seems that the researchers tended to specify scaffolding-related investigations with a more sophisticated manner while they focused on high school level participants. This also to some extent echoes Authors' study (2012) in that a major part of the literature regarding scaffolding gathered empirical data from students in high school contexts (Lin T. C. et al., 2012). Moreover, these researchers also paid attention to the application of computers in scaffolding design (Fang et al., 2016; Dasgupta et al., 2019; Tucker-Raymond et al., 2019; Zheng et al., 2019). The application of computers substantially elaborated the scaffolding that high school learners may require, and hence improved their learning performances in science (Marsteller and Bodzin, 2015; Kern and Crippen, 2017; Moser et al., 2017; Correia et al., 2019; Kyza and Georgiou, 2019).

Co-word Network Analysis Related to Learning Performances

Scaffolding is historically deemed as an effective procedure in instruction that provides learners with assistance or support to perform certain tasks regarding learning. The CNA results hence depict an outline of how researchers notably associated scaffold designs with learning performances (**Figure 8A**). It is anticipated that improvement in learning performance critically guaranteed researchers' meticulous arrangement of scaffolding in their investigations. Almost one-third of the selected articles in this study focused on the co-occurring word "knowledge." This indicates that educators showed remarkable interest in guiding learners' growth of knowledge in light of improvement of academic achievement and conceptual learning. In Fund's (2007) study, the researcher identified effective components of scaffolding in a computerized environment that may positively contribute to junior high school students' learning outcomes. Chen et al. (2016) focused on scaffolding senior high school students' goal setting and planning in learning concepts regarding Boyle's law. This experimental study indicated that male low achievers especially benefitted from the scaffolds built by the researchers. Similar to low achievers, Villanueva et al. (2012) have appealed to science educators to focus their attention on students with special needs. The researchers suggested science writing heuristics as an effective approach that may scaffold students' inquiry, peer-assisted learning, and explanation.

Recently, it has been found that technologies potentially support students' learning and afford additional ways for teacher professional development. Some researchers have proposed a framework to utilize technology for scientific knowledge construction and transfer (Bitan-Friedlander et al., 2004), and understanding of NOS (nature of science) (Sandoval and Reiser, 2004). For example, Ibanez et al. (2016) utilized an image-based augmented reality simulator to scaffold students' science learning. Their findings indicated that proper supports in the secondary school physics curriculum significantly improved students' knowledge about electricity. A review paper that investigated the studies related to mobile applications (apps) and devices for science learning (Zydney and Warner, 2016) found that mobile apps shared some

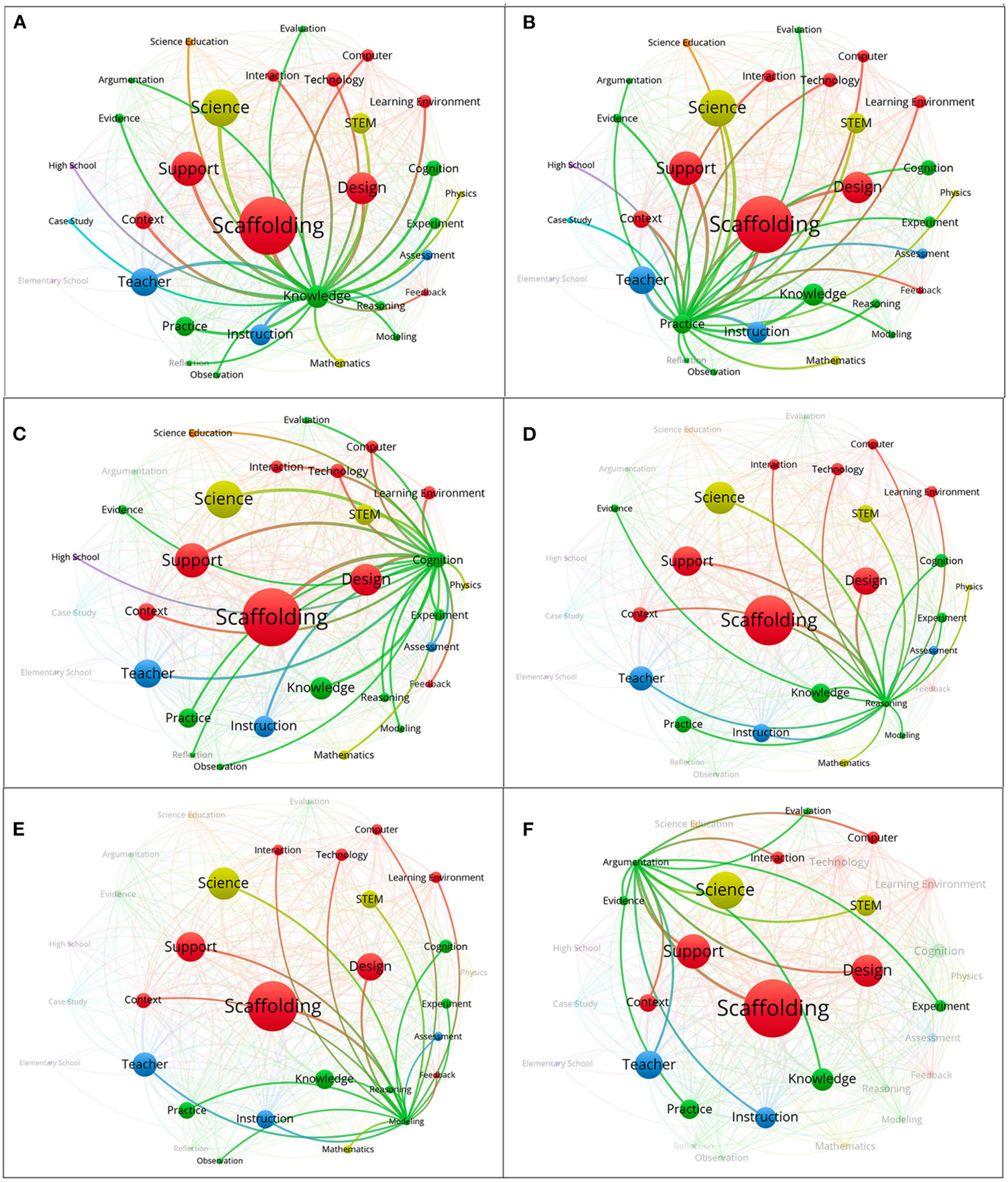


FIGURE 8 | (A) "Knowledge" and other frequently co-word network analysis (keywords = 29). **(B)** "Practice" and other frequently co-word network analysis (keywords = 29). **(C)** "Cognition" and other frequently co-word network analysis (keywords = 27). **(D)** "Reasoning" and other frequently co-word network analysis (keywords = 21). **(E)** "Modeling" and other frequently co-word network analysis (keywords = 21). **(F)** "Argumentation" and other frequently co-word network analysis (keywords = 15).

common design features, cited situated learning theory as the theoretical foundation, and measured scientific knowledge as the most common learning outcomes. Falloon (2017) explored how apps on mobile devices (e.g., iPads) supported hands-on science learning. He also discovered the limitations in the apps to support knowledge development, identified the critical roles of teachers, and clarified the importance of task structure in the promotion of students' conceptual development. Therefore, it is suggested that researchers need to align technologies and learning tools with instructional designs and learning theories for effective supports of students' learning.

Beyond the knowledge perspective, researchers have also tried to reveal possible contributions of scaffolding that benefit learners' learning practice (Fretz et al., 2002; Puente et al., 2013; Fang et al., 2016), cognition in learning (Magana et al., 2012; Zydney et al., 2014; López-Vargas et al., 2017) as well as reasoning (Silk et al., 2009; Kyza et al., 2011; Eggert et al., 2017) (**Figures 8B–D**). For example, Reilly's et al.'s study (2019) demonstrated how thinking scaffolds provided by teachers supported students' reasoning in technology-mediated problem solving and applying experimentation strategies to engage students in scientific reasoning in authentic contexts (Varma, 2014). Applying technological innovation to students' collaborative problem solving (Bressler et al., 2019) and reflection (Berglas-Shapiro et al., 2017) increased in the 5-year period of 2015–2019. As an example of exploring the connections among communications and scientific practices, researchers utilized a mobile game as a collaborative and scaffolding tool to foster collaborative problem solving (Bressler et al., 2019). It is noteworthy that these keywords were found to be overwhelmingly associated with all the key co-occurring words. Such findings hence echo previous reviews of science education literature about increasing academic publications that address learner characteristics (Lin et al., 2014, 2019) and assessing their higher order thinking performances (Zydney and Warner, 2016). There are still remaining research spaces in scaffolding students' development of cognition, reasoning, and scientific practices such as cognitive loads, assessment of higher order thinking performances, and the affordances of technology.

In addition, modeling is prevalent in the selected literature. Modeling is commonly deemed as a complicated mental activity in learning science. Such mental activity requires skills such as analysis, synthesis, and evaluation (VanLehn, 2013). It is not difficult to imagine that learners require certain scaffolds from knowledgeable guides while conducting models (Fretz et al., 2002; Wu et al., 2010; Corpuz and Rebello, 2019). Modeling a natural phenomenon is of value in science learning including explanation of empirical evidence and epistemological shifts when students construct and revise models with technological tools, the teacher's guidance, and peer discussions (Baek and Schwarz, 2015). However, the findings in this study show that the co-occurring relationships between modeling and evidence, modeling and evaluation, as well as modeling and reflection are unexpectedly handful (**Figure 8E**). The reasons for the absent co-occurrence of "modeling and evidence" and "modeling and evaluation" might be attributed to researchers' prevalent interests regarding students' modeling processes in scientific experiments or their mental representations of scientific models. According to the CNA results, studies that simultaneously aim at scaffolding and learners' higher order thinking skills may still require further research efforts.

Compared to the co-word network analysis of modeling, "argumentation" was frequently co-worded with evidence and evaluation, but fewer co-words with limited relationships were uncovered (**Figure 8F**). This reflects a situation contrary to the relevant research trend revealed in previous studies. Lin et al. (2014) and Lin et al. (2019) studies identified argumentation as a frequently published and highly cited issue in the field of science education within the most recent decade. However, while we examined the literature regarding scaffolding, only a handful of studies investigated argumentation at the same time. In the argumentation process, learners not only need the abilities of developing and evaluating arguments based on evidence, but also the ability of examining rebuttals in reflective ways. Moreover, interactivities and milieus in conventional classrooms are additional concerns in developing proper scaffolds for argumentation. To meet the possible difficulties, educators may further consider applying information technologies (Lin H. S. et al., 2012; Belland et al., 2016, 2019) and computer-based argumentation scaffolds (Fretz et al., 2002; Ravenscroft, 2007; Belland et al., 2011) to support learners' argumentation.

CONCLUSIONS AND IMPLICATIONS

The present study reviewed publications in influential literature from 2000 to 2019. Through CNA of the titles, abstracts, and standardized set of keywords of the selected articles, this study unveiled characteristics and trends regarding scaffolding within the recent 20 years. The number of scaffolding-relevant articles shows an obvious growth from 2005–2009 to 2010–2014. In line with Authors' (2012) study, the findings confirmed the increasing trend of scaffolding in both fields of science education and educational technology from various methodological perspectives (Lin T. C. et al., 2012).

Based on the findings from a total of 637 scaffolding-relevant papers, 286 author-defined keywords used by at least two studies were extracted. Among the identified keywords, "scaffolding," "support," and "design" were the top three most frequently used keywords during 2000 and 2019. We hence mapped and concluded six major groups of keywords applied in the selected articles to shape researchers' mutual interest while conducting research relevant to scaffolding. The six groups include scaffolding characteristics, learning performances, teaching, learning disciplines, grades of learners, as well as methodology. Visualization of co-word networks in each 5-year period during 2000 and 2019 revealed an obvious growth of researchers' efforts on "scaffolding characteristics" and "learning performances." More complicated co-word relationships around the keyword "high school" also to some extent reflect both potential and difficulties researchers may encounter while conducting empirical investigations beyond high school contexts. This attempt to integrate co-word network analysis and visualization techniques provides insights regarding certain directions for consecutive investigations. Potentially existing research gaps and niches have also been presented and discussed.

Inevitably, this study may have its own limitations. Regarding the co-word analysis in this study, it is a dilemma to decide which part of each article should be dictionarized. Indeed, failing to dictionarize the full text of all articles may result in a certain degree of miscounting the co-word relationships. It is difficult to conjecture whether or not a word plays a critical role in the main text of the article. In other words, to dictionarize the whole article may strongly diminish the precision of the data analysis. We hence dictionarized only the article titles and abstracts based on the assumption that researchers are inclined to include the most relevant words in these two parts.

AUTHOR CONTRIBUTIONS

T-CL, S-SL, M-LC, and Y-SH contributed to the study conception and design. Material preparation, data collection, and analysis were performed by K-YT and T-CL. Y-SH supervised the research. The first draft of the manuscript was written by all authors. All authors have read and approved the final manuscript.

FUNDING

This work was financially supported by the Institute for Research Excellence in Learning Sciences of National Taiwan Normal University (NTNU) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

REFERENCES

Alrawili, K. S., Osman, K., and Almuntasheri, S. (2020). Effect of scaffolding strategies on higher-order thinking skills in science classroom. *J. Baltic Sci. Educ.* 19, 718–729. doi: 10.33225/jbse/20.19.718

Assefa, S. G., and Rorissa, A. (2013). A bibliometric mapping of the structure of STEM education using co-word analysis. *J. Am. Soc. Inf. Sci. Technol.* 64, 2513–2536. doi: 10.1002/asi.22917

Azevedo, R., Cromley, J. G., and Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemp. Educ. Psychol.* 29, 344–370. doi: 10.1016/j.cedpsych.2003.09.002

*Baek, H., and Schwarz, C. V. (2015). The influence of curriculum, instruction, technology, and social interactions on two fifth-grade students' epistemologies in modeling throughout a model-based curriculum unit. *J. Sci. Educ. Technol.* 24, 216–233. doi: 10.1007/s10956-014-9532-6

*Baker, J. A., Clark, T. P., Maier, K. S., and Viger, S. (2008). The differential influence of instructional context on the academic engagement of students with behavior problems. *Teach. Teach. Educ.* 24, 1876–1883. doi: 10.1016/j.tate.2008.02.019

*Belland, B. R., Glazewski, K. D., and Richardson, J. C. (2011). Problem-based learning and argumentation: testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instruct. Sci.* 39, 667–694. doi: 10.1007/s11251-010-9148-z

*Belland, B. R., Gu, J. Y., Kim, N. J., and Turner, D. J. (2016). An ethnomethodological perspective on how middle school students addressed a water quality problem. *Educ. Technol. Res. Dev.* 64, 1135–1161. doi: 10.1007/s11423-016-9451-8

*Belland, B. R., Walker, A. E., and Kim, N. J. (2017). Synthesizing results from empirical research on computer-based scaffolding in stem education: a meta-analysis. *Rev. Educ. Res.* 87, 309–344. doi: 10.3102/0034654316670999

*Belland, B. R., Weiss, D. M., Kim, N. J., Piland, J., and Gu, J. Y. (2019). An examination of credit recovery students' use of computer-based scaffolding in a problem-based, scientific inquiry unit. *Int. J. Sci. Math. Educ.* 17, 273–293. doi: 10.1007/s10763-017-9872-9

*Berglas-Shapiro, T., Eylon, B. S., and Scherz, Z. (2017). A technology-enhanced intervention for self-regulated learning in science. *Teach. Coll. Rec.* 119, 1–26. doi: 10.1177/016146811711901301

*Bitan-Friedlander, N., Dreyfus, A., and Milgrom, Z. (2004). Types of "teachers in training": the reactions of primary school science teachers when confronted with the task of implementing an innovation. *Teach. Teach. Educ.* 20, 607–619. doi: 10.1016/j.tate.2004.06.007

Bliss, J. (1994). "Children learning science," in *Wonder and Delight: Essays in Science Education in Honour of the Life and Work of Eric Rogers*, eds. B. Jennison and J. Ogborn (Bristol: Institute of Physics Publishing), 45–61.

Bliss, J., Askew, M., and Macrae, S. (1996). Effective teaching and learning: scaffolding revisited. *Oxford Rev. Educ.* 22, 37–61. doi: 10.1080/0305498960220103

*Bressler, D. M., Bodzin, A. M., Eagan, B., and Tabatabai, S. (2019). Using epistemic network analysis to examine discourse and scientific practice during a collaborative game. *J. Sci. Educ. Technol.* 28, 553–566. doi: 10.1007/s10956-019-09786-8

Brownfield, K., and Wilkinson, I. A. G. (2018). Examining the impact of scaffolding on literacy learning: a critical examination of research and guidelines to advance inquiry. *Int. J. Educ. Res.* 90, 177–190. doi: 10.1016/j.ijer.2018.01.004

Callon, M., Courtial, J. P., and Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. *Scientometrics* 22, 155–205. doi: 10.1007/BF02019280

Cambrosio, A., Limoges, C., Courtial, J. P., and Laville, F. (1993). Historical scientometrics? Mapping over 70 years of biological safety research with coword analysis. *Scientometrics* 27, 119–143. doi: 10.1007/BF02016546

Chen, C. K., Huang, N. T., and Hwang, G. J. (in press). Findings and implications of flipped science learning research: a review of journal publications. *Interact. Learn. Environ*. doi: 10.1080/10494820.2019.1690528

*Chen, S. F., Huang, C. C., and Chou, T. L. (2016). The effect of metacognitive scaffolds on low achievers' laboratory learning. *Int. J. Sci. Math. Educ.* 14, 281–296. doi: 10.1007/s10763-015-9691-9

*Choi, A., Klein, V., and Hershberger, S. (2015). Success, difficulty, and instructional strategy to enact an argument-based inquiry approach: experiences of elementary teachers. *Int. J. Sci. Math. Educ.* 13, 991–1011. doi: 10.1007/s10763-014-9525-1

*Corpuz, E. D., and Rebello, N. S. (2019). Refining students' explanations of an unfamiliar physical phenomenon-microscopic friction. *Res. Sci. Educ.* 49, 1177–1211. doi: 10.1007/s11165-017-9650-2

*Correia, A. P., Koehler, N., Thompson, A., and Phye, G. (2019). The application of PhET simulation to teach gas behavior on the submicroscopic level: secondary school students' perceptions. *Res. Sci. Technol. Educ.* 37, 193–217. doi: 10.1080/02635143.2018.1487834

*Dasgupta, C., Magana, A. J., and Vieira, C. (2019). Investigating the affordances of a CAD enabled learning environment for promoting integrated STEM learning. *Comput. Educ.* 129, 122–142. doi: 10.1016/j.compedu.2018.10.014

*Davis, E. A., and Linn, M. C. (2000). Scaffolding students' knowledge integration: prompts for reflection in KIE. *Int. J. Sci. Educ.* 22, 819–837. doi: 10.1080/095006900412293

*Decristan, J., Hondrich, A. L., Buttner, G., Hertel, S., Klieme, E., Kunter, M., et al. (2015). Impact of additional guidance in science education on primary students' conceptual understanding. *J. Educ. Res.* 108, 358–370. doi: 10.1080/00220671.2014.899957

Dehdarirad, T., Villarroya, A., and Barrios, M. (2014). Research trends in gender differences in higher education and science: a co-word analysis. *Scientometrics* 101, 273–290. doi: 10.1007/s11192-014-1327-2

*Eggert, S., Nitsch, A., Boone, W. J., Nuckles, M., and Bogeholz, S. (2017). Supporting students' learning and socioscientific reasoning about climate change-the effect of computer-based concept mapping scaffolds. *Res. Sci. Educ.* 47, 137–159. doi: 10.1007/s11165-015-9493-7

*Falloon, G. (2017). Mobile devices and apps as scaffolds to science learning in the primary classroom. *J. Sci. Educ. Technol.* 26, 613–628. doi: 10.1007/s10956-017-9702-4

Fang, S. C., Hsu, Y. S., and Hsu, W. H. (2016). Effects of explicit and implicit prompts on students' inquiry practices in computer-supported learning environments in high school earth science. *Int. J. Sci. Educ.* 38, 1699–1726. doi: 10.1080/09500693.2016.1213458

*Fretz, E. B., Wu, H. K., Zhang, B. H., Davis, E. A., Krajcik, J. S., and Soloway, E. (2002). An investigation of software scaffolds supporting modeling practices. *Res. Sci. Educ.* 32, 567–589. doi: 10.1023/A:1022400817926

*Fund, Z. (2007). The effects of scaffolded computerized science problem-solving on achievement outcomes: a comparative study of support programs. *J. Comput. Assist. Learn.* 23, 410–424. doi: 10.1111/j.1365-2729.2007.00226.x

Hsu, Y. S., Lai, T. L., and Hsu, W. H. (2015). A design model of distributed scaffolding for inquiry-based learning. *Res. Sci. Educ.* 45, 241–273. doi: 10.1007/s11165-014-9421-2

Huang, C., Yang, C., Wang, S., Wu, W., Su, J., and Liang, C. (2020). Evolution of topics in education research: a systematic review using bibliometric analysis. *Educ. Rev.* 72, 281–297. doi: 10.1080/00131911.2019.1566212

*Ibanez, M. B., Di-Serio, A., Villaran-Molina, D., and Delgado-Kloos, C. (2016). Support for agmented reality simulation systems: the effects of scaffolding on learning outcomes and behavior patterns. *IEEE Trans. Learn. Technol.* 9, 46–56. doi: 10.1109/TLT.2015.2445761

Kelley, T. L. (1927). *Interpretation of Educational Measurements*. New York, NY: Macmillan.

*Kern, C. L., and Crippen, K. J. (2017). The effect of scaffolding strategies for inscriptions and argumentation in a science cyberlearning environment. *J. Sci. Educ. Technol.* 26, 33–43. doi: 10.1007/s10956-016-9649-x

*Kershner, R., Mercer, N., Warwick, P., and Staarman, J. K. (2010). Can the interactive whiteboard support young children's collaborative communication and thinking in classroom science activities? *Int. J. Comput. Support. Collab. Learn.* 5, 359–383. doi: 10.1007/s11412-010-9096-2

*Kim, M. C., and Hannafin, M. J. (2011). Scaffolding problem solving in technology-enhanced learning environments (TELEs): bridging research and theory with practice. *Comput. Educ.* 56, 403–417. doi: 10.1016/j.compedu.2010.08.024

*Kyza, E. A., Constantinou, C. P., and Spanoudis, G. (2011). Sixth graders' co-construction of explanations of a disturbance in an ecosystem: exploring relationships between grouping, reflective scaffolding, and evidence-based explanations. *Int. J. Sci. Educ.* 33, 2489–2525. doi: 10.1080/09500693.2010.550951

*Kyza, E. A., and Georgiou, Y. (2019). Scaffolding augmented reality inquiry learning: the design and investigation of the TraceReaders location-based, augmented reality platform. *Interact. Learn. Environ.* 27, 211–225. doi: 10.1080/10494820.2018.1458039

*Land, S. M., and Zembal-Saul, C. (2003). Scaffolding reflection and articulation of scientific explanations in a data-rich, project-based learning environment: an investigation of progress portfolio. *Educ. Technol. Res. Dev.* 51, 65–84. doi: 10.1007/BF02504544

Leydesdorff, L. (1989). The relations between qualitative theory and scientometric methods in science and technology studies. *Scientometrics* 15, 333–347. doi: 10.1007/BF02017058

*Lin, H. S., Hong, Z. R., and Lawrenz, F. (2012). Promoting and scaffolding argumentation through reflective asynchronous discussions. *Comput. Educ.* 59, 378–384. doi: 10.1016/j.compedu.2012.01.019

*Lin, T. C., Hsu, Y. S., Lin, S. S., Changlai, M. L., Yang, K. Y., and Lai, T. L. (2012). A review of empirical evidence on scaffolding for science education. *Int. J. Sci. Math. Educ.* 10, 437–455. doi: 10.1007/s10763-011-9322-z

Lin, T. C., Lin, T. J., and Tsai, C. C. (2014). Research trends in science education from 2008 to 2012: A systematic content analysis of publications in selected journals. *Int. J. Sci. Educ.* 36, 1346–1372. doi: 10.1080/09500693.2013.864428

Lin, T. J., Lin, T. C., Potvin, P., and Tsai, C. C. (2019). Research trends in science education from 2013 to 2017: A systematic content analysis of publications in selected journals. *Int. J. Sci. Educ.* 41, 367–387. doi: 10.1080/09500693.2018.1550274

López-Belmonte, J., Moreno-Guerrero, A. J., Pozo-Sánchez, S., and Marín-Marín, J. A. (2021). Co-word analysis and academic performance from the Australasian Journal of Educational Technology in Web of Science. *Aust. J. Educ. Technol.* 37, 119–140. doi: 10.14742/ajet.6940

*López-Vargas, O., Ibáñez-Ibáñez, J., and Racines-Prada, O. (2017). Students' metacognition and cognitive style and their effect on cognitive load and learning achievement. *Educ. Technol. Soc.* 20, 145–137. doi: 10.1177/1365480217704263

*Magana, A. J., Brophy, S. P., and Bryan, L. A. (2012). An integrated knowledge framework to characterize and scaffold size and scale cognition (FS2C). *Int. J. Sci. Educ.* 34, 2181–2203. doi: 10.1080/09500693.2012.715316

Marín-Marín, J. A., Moreno-Guerrero, A. J., Dúo-Terrón, P., and López-Belmonte, J. (2021). STEAM in education: a bibliometric analysis of performance and co-words in Web of Science. *Int. J. STEM Educ.* 8, 41. doi: 10.1186/s40594-021-00296-x

*Marsteller, R. B., and Bodzin, A. M. (2015). The effectiveness of an online curriculum on high school students' understanding of biological evolution. *J. Sci. Educ. Technol.* 24, 803–817. doi: 10.1007/s10956-015-9565-5

*McNeill, K. L., Lizotte, D. J., Krajcik, J., and Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *J. Learn. Sci.* 15, 153–191. doi: 10.1207/s15327809jls1502_1

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. *Rev. Educ. Res.* 65, 93–127. doi: 10.3102/00346543065002093

Metz, K. E. (1997). On the complex relation between cognitive developmental research and children's science curricula. *Rev. Educ. Res.* 67, 151–163. doi: 10.3102/00346543067001151

*Moser, S., Zumbach, J., and Deibl, I. (2017). The effect of metacognitive training and prompting on learning success in simulation-based physics learning. *Sci. Educ.* 101, 944–967. doi: 10.1002/sce.21295

Noroozi, O., Kirschner, P. A., Biemans, H. J., and Mulder, M. (2018). Promoting argumentation competence: extending from first- to second-order scaffolding through adaptive fading. *Educ. Psychol. Rev.* 30, 153–176. doi: 10.1007/s10648-017-9400-z

Oh, S., and Jonassen, D. H. (2007). Scaffolding online argumentation during problem solving. *J. Comput. Assist. Learn.* 23, 95–110. doi: 10.1111/j.1365-2729.2006.00206.x

*Puente, S. M. G., van Eijck, M., and Jochems, W. (2013). A sampled literature review of design-based learning approaches: a search for key characteristics. *Int. J. Technol. Des. Educ.* 23, 717–732. doi: 10.1007/s10798-012-9212-x

*Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., et al. (2004). A scaffolding design framework for software to support science inquiry. *J. Learn. Sci.* 13, 337–386. doi: 10.1207/s15327809jls1303_4

*Ravenscroft, A. (2007). Promoting thinking and conceptual change with digital dialogue games. *J. Comput. Assist. Learn.* 23, 453–465. doi: 10.1111/j.1365-2729.2007.00232.x

*Reilly, C. M., Kang, S. Y., Grotzer, T. A., Joyal, J. A., and Oriol, N. E. (2019). Pedagogical moves and student thinking in technology-mediated medical problem-based learning: supporting novice-expert shift. *Br. J. Educ. Technol.* 50, 2234–2250. doi: 10.1111/bjet.12843

*Reiser, B. J. (2004). Scaffolding complex learning: the mechanisms of structuring and problematizing student work. *J. Learn. Sci.* 13, 273–304. doi: 10.1207/s15327809jls1303_2

Rogoff, B. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. Oxford: Oxford University Press.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., and Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Soc. Work Res.* 27, 94–104. doi: 10.1093/swr/27.2.94

*Sandoval, W. A., and Reiser, B. J. (2004). Explanation-driven inquiry: integrating conceptual and epistemic scaffolds for scientific inquiry. *Sci. Educ.* 88, 345–372. doi: 10.1002/sce.10130

*Silk, E. M., Schunn, C. D., and Cary, M. S. (2009). The impact of an engineering design curriculum on science reasoning in an urban setting. *J. Sci. Educ. Technol.* 18, 209–223. doi: 10.1007/s10956-009-9144-8

Tang, K. T., Wang, C. Y., Chang, H. Y., Chen, S. F., Lo, H. C., and Tsai, C. C. (2016). The intellectual structure of metacognitive scaffolding in science education: A co-citation network analysis. *Int. J. Sci. Math. Educ.* 14, 249–262. doi: 10.1007/s10763-015-9696-4

Toleda, S., and Dubas, J. M. (2016). Encouraging higher-order thinking in general chemistry by scaffolding student learning using Marzano's taxonomy. *J. Chem. Educ.* 93, 64–69. doi: 10.1021/acs.jchemed.5b00184

*Tucker-Raymond, E., Puttick, G., Cassidy, M., Harteveld, C., and Troiano, G. M. (2019). "I Broke Your Game!": critique among middle schoolers designing computer games about climate change. *Int. J. STEM Educ.* 6:41. doi: 10.1186/s40594-019-0194-z

*van Uum, M. S. J., Verhoeff, R. P., and Peeters, M. (2017). Inquiry-based science education: Scaffolding pupils' self-directed learning in open inquiry. *Int. J. Sci. Educ.* 39, 2461–2481. doi: 10.1080/09500693.2017.1388940

*VanLehn, K. (2013). Model construction as a learning activity: a design space and review. *Interact. Learn. Environ.* 21, 371–413. doi: 10.1080/10494820.2013.803125

*Varma, K. (2014). Supporting scientific experimentation and reasoning in young elementary school students. *J. Sci. Educ. Technol.* 23, 381–397. doi: 10.1007/s10956-013-9470-8

*Villanueva, M. G., Taylor, J., Therrien, W., and Hand, B. (2012). Science education for students with special needs. *Stud. Sci. Educ.* 48, 187–215. doi: 10.1080/14703297.2012.737117

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wood, D., Bruner, J. S., and Ross, G. (1976). The role of tutoring in problem solving. *J. Child Psychol. Psychiatry Allied Discipl.* 17, 89–100. doi: 10.1111/j.1469-7610.1976.tb00381.x

Wu, H. K., Hsu, Y. S., and Hwang, F. K. (2010). Designing a technology-enhanced learning environment to support scientific modeling. *Turkish Online J. Educ. Technol.* 9, 58–65.

Wu, S. H., Lai, C. L., Hwang, G. J., and Tsai, C. C. (2021). Research trends in technology-enhanced chemistry learning: a review of comparative research from 2010 to 2019. *J. Sci. Educ. Technol.* 30, 496–510. doi: 10.1007/s10956-020-09894-w

*Zheng, J., Xing, W., and Zhu, G. X. (2019). Examining sequential patterns of self- and socially shared regulation of STEM learning in a CSCL environment. *Comput. Educ.* 136, 34–48. doi: 10.1016/j.compedu.2019.03.005

Zhou, M., and Lam, K. K. (2019). Metacognitive scaffolding for online information search in K-12 and higher education settings: a systematic review. *Educ. Technol. Res. Dev.* 67, 1353–1384. doi: 10.1007/s11423-019-09646-7

*Zydney, J. M., Bathke, A., and Hasselbring, T. S. (2014). Finding the optimal guidance for enhancing anchored instruction. *Interact. Learn. Environ.* 22, 668–683. doi: 10.1080/10494820.2012.745436

*Zydney, J. M., and Warner, Z. (2016). Mobile apps for science learning: review of research. *Comput. Educ.* 94, 1–17. doi: 10.1016/j.compedu.2015.11.001

*References included in research samples are asterisked.

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

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

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fpsyg.2022.875161

Development and Testing of the Curiosity in Classrooms Framework and Coding Protocol

Jamie J. Jirout[1]*, Sharon Zumbrunn[2], Natalie S. Evans[1] and Virginia E. Vitiello[1]

[1]School of Education and Human Development, University of Virginia, Charlottesville, VA, United States, [2]School of Education, Virginia Commonwealth University, Richmond, VA, United States

Curiosity is widely acknowledged as a crucial aspect of children's development and as an important part of the learning process, with prior research showing associations between curiosity and achievement. Despite this evidence, there is little research on the development of curiosity or on promoting curiosity in school settings, and measures of curiosity promotion in the classroom are absent from the published literature. This article introduces the Curiosity in Classrooms (CiC) Framework coding protocol, a tool for observing and coding instructional practices that support the promotion of curiosity. We describe the development of the framework and observation instrument and the results of a feasibility study using the protocol, which gives a descriptive overview of curiosity-promoting instruction in 35 elementary-level math lessons. Our discussion includes lessons learned from this work and suggestions for future research using the developed observation tool.

Keywords: curiosity, education, observations, instruction, protocol development

Edited by:
Kai-Yu Tang,
Ming Chuan University, Taiwan

Reviewed by:
Tzu-Chiang Lin,
National Kaohsiung University of Science and Technology, Taiwan
Te-Lien Chou,
Fu Jen Catholic University, Taiwan

****Correspondence:***
Jamie J. Jirout
jirout@virginia.edu

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

Received: *13 February 2022*
Accepted: *15 March 2022*
Published: *07 April 2022*

Citation:
Jirout JJ, Zumbrunn S, Evans NS and Vitiello VE (2022) Development and Testing of the Curiosity in Classrooms Framework and Coding Protocol. Front. Psychol. 13:875161. doi: 10.3389/fpsyg.2022.875161

INTRODUCTION

> Curiosity needs food as much as any of us, and dies soon if denied it.
> -Stella Benson (*I Pose*)

Curiosity is widely acknowledged as a crucial aspect of children's development, and as an important part of the learning process (Jirout and Klahr, 2012). Evidence suggests associations between curiosity and achievement at school entry (Shah et al., 2018) and that curiosity supports academic performance, even when controlling for students' effort and ability (von Stumm et al., 2011). Despite this evidence, most prior research on the development of curiosity or on promoting curiosity has been conducted in lab settings with individual children (e.g., Cook et al., 2011; Gweon et al., 2014; Shneidman et al., 2016; Danovitch et al., 2021 among others), rather than in school settings. In research that did look at promoting curiosity in an educational context, researchers test specific manipulations with researchers administering the lesson to promote curiosity in schools (e.g., Lamnina and Chase, 2019) or parents in a museum setting (e.g., Willard et al., 2019), or observed children's exploration without studying instruction or promotion of curiosity (e.g., van Schijndel et al., 2018). To our knowledge, measures of curiosity promotion in the classroom are absent from the published literature. We extend this prior work by focusing on more general practices that can be used to promote student curiosity in classroom contexts, developing a protocol for measuring this promotion, and showing its feasibility of use for future research.

There is concern that curiosity declines with education (Coie, 1974; Engelhard and Monsaas, 1988; Engel, 2015), so it is important to identify and provide what Benson refers to as the "food" for curiosity in educational contexts. In this methodological piece, we introduce the Curiosity in Classrooms (CiC) Framework coding protocol, a tool for observing and coding instructional practices that support the promotion of curiosity (Jirout et al., 2018). We describe the development of the observation instrument and the results of a feasibility study using the protocol with the intent to make the instrument available to the research and evaluation community.

Defining Curiosity

A challenge in studying curiosity in education begins with the initial challenge of defining and operationalizing the construct of curiosity. Curiosity can be described as multidimensional, with theories suggesting different "types" or dimensions of curiosity (Kidd and Hayden, 2015), but it is also multifaceted in that it can include affective, cognitive, motivational, physiological, and expressive processes (Jirout and Klahr, 2012). For example, curiosity has been described as recognition that there is something unknown that one wants to know, or that there is ambiguity or uncertainty to resolve (cognitive; Jirout and Klahr, 2012), excitement or anticipation of pleasure from learning something new, or uneasiness of not knowing something (affective; Litman et al., 2005; Litman, 2008), and the desire to seek information by exploring or asking questions (motivational; Ryan, 2012; Pekrun and Linnenbrink-Garcia, 2012), all of which can be important influences in educational contexts.

Here we operationalize curiosity as a desire to seek information when something is unknown, and especially a preference to explore and gather information under conditions of uncertainty (Loewenstein, 1994; Jirout and Klahr, 2012). This definition stems from research showing that uncertainty leads to greater levels of exploration, with less exploration when there is too little or too much uncertainty (Loewenstein, 1994; Litman et al., 2005), and individual differences in uncertainty preferences, or the levels that an individual considers to be too little or too much (Jirout and Klahr, 2012). This assumption considers both a state aspect of curiosity, in that curiosity is momentarily sparked in response to the presence of uncertainty, as well as a more stable aspect of curiosity in the difference in preferred uncertainty for exploration, which can lead to higher likelihood to have curiosity sparked across contexts more generally. Rather than trying to disentangle the debate about curiosity being a state or a trait, or how it should be considered in both ways in education research (e.g., Engel, 2015; Murayama et al., 2019), we suggest that more regular experience of curiosity as a state can also lead to developing more stable curiosity, discussed further below. If curiosity includes both the desire for information and exploration to gather that information, regular promotion of curiosity in classrooms would result in more frequent feelings of curiosity and information-seeking behavior. At a more stable level of curiosity (i.e., individual differences), promoting curiosity would lead to greater comfort with uncertainty over time, so that preferences for exploring would relate to greater levels of information gaps (Jirout and Klahr, 2012; Jirout et al., 2018).

Curiosity, Learning, and Education

In general, curiosity is associated with motivation and behavior that is conducive for learning, such as engagement and persistence in facing obstacles and setting goals (Kashdan and Steger, 2007), in developing sustained interests, which, in turn, can promote self-regulation, information-seeking, and motivation (Renninger, 2000; Hidi and Renninger, 2006), and with social, emotional, and cognitive development across the lifespan more generally (Kashdan, 2006; Kashdan and Steger, 2007; Keller et al., 2012). Aligned with associations between curiosity and learning, teachers have a positive perception of student curiosity and view its role in learning as distinct from traits, such as creativity and imagination (Chak, 2007). Further, teachers consider curiosity to be a didactic tool in the classroom, supporting better relationships with peers during group activities, encouraging critical thinking, and fostering feelings of self-determination (Menning, 2019). Teachers rate more curious children higher in competence motivation, attention, and persistence (Jirout and Klahr, 2012), and children with higher levels of curiosity are generally perceived by teachers as more likely to explore, share their interests with peers, and express excitement (Spektor-Levy et al., 2013).

Despite this extensive valuing and benefit of curiosity for learning, and as researchers and educators lament, children's curiosity seems to diminish as they progress through formal education, at least curiosity expressed in school (Engelhard and Monsaas, 1988; Engel, 2011; Post and Walma van der Molen, 2018). Levels of expressed curiosity in students are very low by 1st grade, and almost completely absent by 5th grade (Engel and Randall, 2009). While research explaining this pattern is lacking, one suggested explanation of low curiosity in schools is an inconsistency between the emphasis on performance in schools and student curiosity (Engel, 2015; Jirout et al., 2018). For example, traditional instructional assignments often focus on getting correct answers or doing things the "right" way, leaving little room for students to question, wonder, or try out new or different ways of doing things. If this is true, it might be that children are not losing curiosity, but rather simply are not curious while in school. Such an argument is consistent with studies comparing children's curiosity within and outside of school (Tizard and Hughes, 1984; Post and Walma van der Molen, 2018). For example, Tizard et al. (1983) assessed preschool children's questions at home and in school and found that children asked more than ten times as many questions at home, with hourly rates of "curiosity" questions at home vs. school as 2.3 vs. 20.6 and 0.3 vs. 12.3 for middle- and working-class samples, respectively.

Why might children ask fewer questions and be less curious in school? In a recent study that surveyed preschool and elementary-aged children about what they were curious about, children rarely responded with curiosities related to school spontaneously, and, when prompted about what they were curious about in school, responses were mostly unrelated to investigative learning (Post and Walma van der Molen, 2018). Results also showed inconsistency between being curious and children's perceptions of expectations in school in some reported responses. In the report, the researchers reported some children as "quite surprised or even disturbed when we asked them to share their school-specific curiosities," with example responses of: "No one is curious about what we learn in class. We just need to do whatever the teachers tell us to do," "It does not matter whether I am curious, because we just need to learn whatever we are assigned to do," and "Are you joking? There is nothing to be curious about, when doing boring math or reading" (Post and Walma van der Molen, 2018, p. 65). With such significance for lifelong learning and wellbeing, it is important to better understand how curiosity is manifested in formal educational contexts and how curiosity can be promoted through instructional practices.

Promoting Curiosity in Education

To explore the ways in which educational contexts can promote curiosity, it is important to first identify what it looks like to be curious in classrooms, and what that means for what a change in curiosity might look like. Based on the operationalization described earlier, we suggest that being curious in a classroom means that students recognize and feel comfortable with uncertainty, which leads to them becoming curious and engaging in exploration to find the missing information. Importantly, for students to feel comfortable being curious in a classroom setting, we hypothesize that they must have the perception that their teacher welcomes and values curiosity, and believe that curiosity is important for learning and has a place in school (Post and Walma van der Molen, 2018).

Our operationalization of curiosity as information-seeking in response to knowledge gaps (Loewenstein, 1994; Jirout and Klahr, 2012) describes the cognitive processes involved in curiosity and provides direction for identifying potential methods of promoting curiosity in education, including both individual student characteristics and contextual factors. When specifically considering individual factors, being more curious means having a higher preference or threshold for uncertainty, that is, being more likely to explore and seek information in the presence of larger knowledge gaps. Prior research suggests that the relation between uncertainty and curiosity follows an inverted U shape curve, where there is an optimal level of uncertainty that will lead to curiosity and exploration (Loewenstein, 1994; Jirout and Klahr, 2012; Kidd et al., 2012). For example, several studies measured children's curiosity using a game in which children chose what they wanted to explore, with the choice between options that gave more information or less information (i.e., introduced more uncertainty) about what they would find; choosing to explore the option that presented more uncertainty indicated higher curiosity (Jirout and Klahr, 2012; Gordon et al., 2015; van Schijndel et al., 2018). Children showed individual differences in the level of uncertainty that led to exploration, and these differences were associated with convergent measures, such as question asking behavior and teacher-rated learning behaviors (e.g., competence motivation and persistence, Jirout and Klahr, 2012). Translating to an educational context, this might look like a student choosing a project they know less about over one in which they already know much of the information, or a student reading beyond what is assigned because they want to know more even after they have read the information needed to do an assignment.

We suggest that continuously promoting curiosity during instruction in ways that are positively experienced can help children to develop a more stable comfort with uncertainty, thus positively influencing their more stable curiosity over time. This is similar to the concept of the Broaden and Build theory (Fredrickson, 1998), which suggests that experiencing positive affect for learning can lead to an "upward spiral" of broadened cognition that further supports future positive affect (Fredrickson and Joiner, 2002) and to the reward-learning framework of curiosity and interest (Murayama et al., 2019). Murayama and colleagues suggest that engaging in momentary information-seeking can support future information-seeking more generally by reinforcing the behavior through the "reward" of gaining the valued information (i.e., an intrinsically rewarding feedback loop), as well as by expanding one's knowledge base, which can lead to new questions and new opportunities to explore (Murayama et al., 2019; Murayama, 2022). This differs from what we theorize to occur in that it focuses on changes based on reinforcement through rewards. Our conceptualization of curiosity focuses on changes in the amount of uncertainty learners prefer. In other words, the reward-learning framework describes how the reward feedback loop promotes increased engagement in information-seeking when there is uncertainty they want to resolve, while we focus on learners' decisions about what to explore, with possible options often varying across levels of uncertainty. Rather than seeing increases in general frequency of information-seeking with increasing curiosity, we would expect to see increases in the amount of uncertainty learners want to explore—whether they choose things that are more or less uncertain, and, thus, result in more or less that can be learned.

In our earlier work, we developed a framework identifying the means by which teachers can create a classroom climate that promotes students' curiosity through instruction and language, as well as ways curiosity might be suppressed: The Curiosity in Classrooms (CiC) framework (Jirout et al., 2018). The CiC framework suggests two ways by which instruction might promote curiosity: (1) helping students to recognize and become more comfortable with uncertainty (initiating or sparking curiosity), and (2) helping children learn to seek information to resolve curiosity (promoting curious behavior—that is, exploration and information-seeking). Importantly, the CiC framework illustrates how the instructional language teachers use to present content and learning activities might promote or suppress curiosity. This approach aligns with the process–product model of Brophy and Good (1984), which has been successful in prior research identifying effective contextual factors and teaching practices. For example, studies using this model have shown that classroom climate, and specifically teacher interactions, can create a nurturing and supportive environment for students to learn (Hamre and

TABLE 1 | Curiosity in classrooms framework: types and examples of curiosity-promoting instruction.

Curiosity promotion		
Categories	**Practice**	**Example**
C1: Promoting feelings of curiosity (recognition of and desire to explore uncertainty)	C1.1 Opportunities to think, question, and participate	"Take a few minutes to look at the image, and think about what you notice or wonder."
	C1.2 Modeling positive reactions to uncertainty	"You know, sometimes I get confused, too."
	C1.3 Prompting question generation	"Who can share questions we could ask to learn more about this?"
	C1.4 Reviewing known and unknown information and making connections	"We know that alligators are reptiles. What do we know about reptiles? What might that tell us about how alligators live?"
	C1.5 Encouraging alternative ideas	"Who did something different—can someone share another way we could try to solve this problem?"
C2: Promoting curious behaviors (exploration and questioning to resolve uncertainty)	C2.1 Opportunities to explore ideas, materials, and questions	"Now that you have the cubes, try and use them to explore different ways to show fractions."
	C2.2 Scaffolding information-seeking	"I bet you can find that out—what could we search for on Google that might have some information?"
	C2.3 Positive verbal and non-verbal responses to students' questions	"What an interesting question!"
Curiosity Suppression		
S1: Avoiding uncertainty and promoting discomfort with uncertainty		"I'm not sure why it looks different but we need to move on, so just pay attention to the picture in your book for what it should look like."
S2: Actively discouraging information-seeking behaviors		"Your materials do not look like you are following the instructions; stop playing around and focus on the question you are supposed to answer."

Pianta, 2005). Curiosity and learning beliefs are likely to be highly sensitive to specific language used, and, importantly, these factors are malleable. For example, the growing number of interventions related to the construct of mindset demonstrates its plasticity, and teacher language could be an important influence of developing mindset beliefs about learning (Yeager et al., 2013). For example, children who learned from a robot that modeled a growth mindset during instruction, compared to a control robot that did not model growth mindset, agreed more strongly with growth mindset statements (H. Park et al., 2017). Similarly, children who learned from a "curious" robot teacher—one that asked probing questions—scored higher on curiosity and exploration tasks than those whose robot teacher gave the same information without questions (Gordon et al., 2015). These studies show the effect of language can have on curiosity and learning attitudes.

The CiC framework of instructional practices to promote curiosity is presented with examples in **Table 1**, and a detailed explanation of prior theoretical and empirical literature supporting the different framework components can be found in our past work (see Jirout et al., 2018). However, the framework development was influential in describing the development of our observational tool and is not described elsewhere, so in this work, we begin by describing the approach we used to identify specific instructional practices that can promote curiosity. We then focus on the primary goal of this work: to develop a coding protocol for and test of the framework.

MATERIALS AND METHODS

Design and Development Process

The CiC framework is a set of instructional practices identified to promote curiosity, while the CiC framework coding protocol is a guide describing how to identify practices observed during instruction with detailed specifications for what is and what is not included within each of the framework elements. The CiC framework and coding protocol development involved two stages with several steps within each to identify and measure instances in which curiosity could be promoted or suppressed by teachers (see **Figure 1**).

Stage 1. Establishing the CiC Conceptual Framework

The development of the CiC framework and identification of instructional practices that either promote or suppress curiosity began with an empirical literature review (see Jirout et al., 2018) and observations of video-recordings of classroom instruction from prior unrelated studies and from online resources, such as teachingchannel.org. A multi-disciplinary team of researchers with expertise in studying motivation, curiosity, effective K-12 pedagogy, and/or developmental psychology and in developing observational tools met across multiple sessions to watch and discuss instruction with the goal of observing behaviors relating to the theoretically identified categories of curiosity promotion and suppression based on the Information-Gap Theory (Loewenstein, 1994). From these observations, many behaviors with the potential to positively or negatively influence student curiosity were identified. These were then grouped by theme, resulting in the promotion and suppression categories (see **Table 1**).

In describing what curiosity promotion might look like in classrooms, we focused on specific ways that students might become more comfortable with uncertainty, practice recognizing uncertainty, and learn to and practice resolving uncertainty, and the corresponding teacher supportive behaviors for each of these categories. As such, we divided curiosity promotion into two categories: promoting feelings of curiosity (C1) and promoting curious behaviors (C2), aligned with eight practice

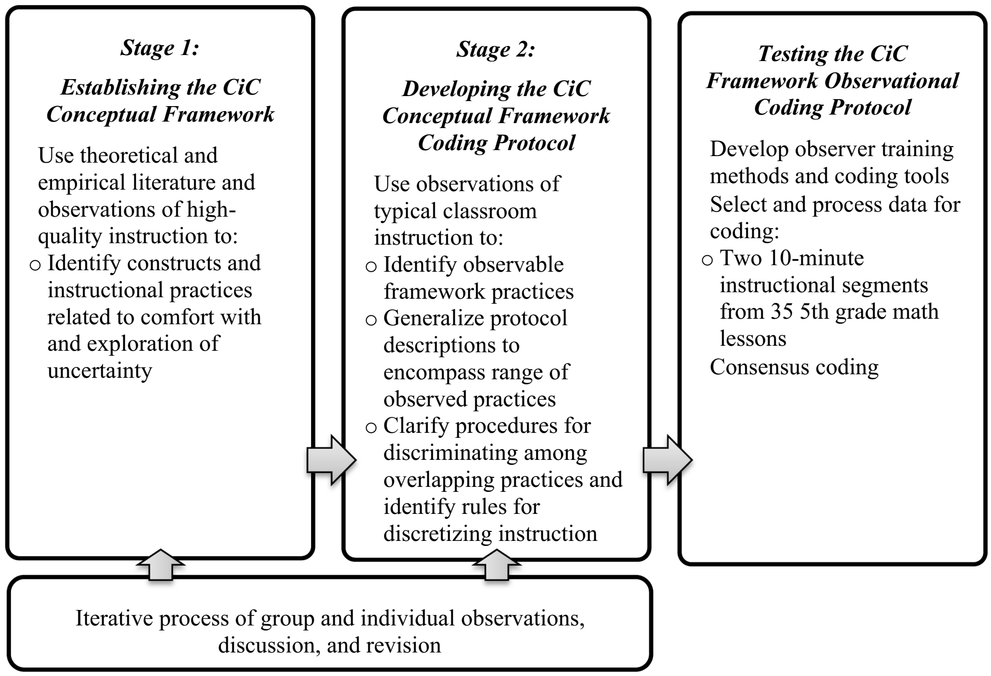


FIGURE 1 | Overview of the CiC framework and protocol development process.

types across the two categories. We also identified two categories of curiosity suppression: promoting discomfort with or avoiding uncertainty, and actively discouraging information-seeking behaviors, each aligned with one practice type. In the following sections, we describe categories and associated practices in more detail.

C1. Promoting Feelings of Curiosity

The category of *promoting feelings of curiosity* addresses teacher actions and utterances that create an environment supportive of students' curiosity. This is done by promoting a climate where students feel safe taking risks, making mistakes, failing, and not knowing or not being sure (Dess and Picken, 2012; Jirout et al., 2018). Teachers' behavior and language can promote this type of classroom climate by providing opportunity and encouragement for these behaviors, as well as modeling the behaviors in themselves (Zimmerman and Pike, 1972; Sullivan, 2011). This category looks specifically at how the teacher prompts students to be curious and guides children on how to be curious. Actions and utterances within this category promote curiosity (and children's learning to be curious) rather than promoting and developing skills involved in resolving curiosity. Specifically, five instructional practices were identified as methods of promoting students' ability to recognize uncertainty or to develop greater comfort with it.

First, to promote feelings of curiosity, teachers can provide students with opportunities to think, question, participate, and respond, such as having students take a moment to think about something and share it with a peer before having a student respond to the whole class, allowing broader participation and engagement (Rowe, 1972). For example, asking questions that guide children's thinking rather than giving them information is found to support exploration and broader learning (Bonawitz et al., 2011; Yu et al., 2018; Jean et al., 2019). Providing opportunities for students to become curious can promote their engagement in learning, as suggested in a study where students were more likely to explore hints about information they were curious to know then to just get the answer (Metcalfe et al., 2021). Related to this, a second approach teachers can take is to elicit multiple responses from several students, allowing them to hear each other's perspectives and ideas and showing that there are different ways of thinking about things (Duschl and Osborne, 2002). For example, students instructed to seek differing ideas from other students in a science class learned more than students who were asked to seek ideas from classmates that matched their own (Matuk and Linn, 2015). Common activities, such as Think-Pair-Share, can support these opportunities for children to become curious, active participants in their learning (Rowe, 1986; McTighe and Lyman, 1988; Ciardiello, 1998), and simply waiting a few seconds longer after asking questions can increase the number, types, and quality of student responses (Rowe, 1972). Third, teachers can model their own comfort with uncertainty or mistakes, showing a sense of wonder about something they do not know, or pointing out the benefit of learning something new when exploring what led to a mistake (Engel, 2015), which can also broaden their own thinking about the material (Zimmerman and Pike, 1972; McDonald, 1992). When teachers focus more on knowing how to find information and are comfortable acknowledging the limits to their knowledge, they are more open to student interests, and students have deeper and more meaningful questions and discussions (Cunnigham, 1997).

Benefits of a "motivation culture" in classrooms are found for both motivation and achievement, and stronger perceptions of mistakes as useful for learning are associated with higher motivation (Käfer et al., 2019). Fourth, teachers can help to support the metacognitive process of considering what is known related to a question, and what is unknown or what information is needed, helping children to explicitly recognize knowledge gaps (King, 1994; Rosenshine et al., 1996). Fifth, students can be explicitly prompted to generate questions, different from simply asking "are there any questions?" but rather giving students the task of thinking of what questions they could ask to learn more (e.g., "what questions could we ask to learn more about this?"). Having students' own questions included in the learning process can support a sense of autonomy for learning, which is a powerful motivator for learning (Grolnick and Ryan, 1987). Together, these methods can all help to support a learning culture where mistakes are not seen as something to try and avoid (Martin and Marsh, 2003; Martin, 2011), and instead puts the focus on the learning process whereby inquiry is valued and appreciated.

C2. Promoting Curious Behaviors

The second category of curiosity promotion, *promoting curious behaviors*, includes language and behavior intended to provide the opportunities, instructions, and reinforcement of manifestations of curiosity, (i.e., curious behaviors including information-seeking, such as exploration and asking questions). This category includes teacher actions and utterances encouraging students to act on their curiosity, as a result of either the classroom settings/provocations, teacher verbal and non-verbal responses/prompts, and/or explicit communications about how to be curious. A key distinction between the C1 and C2 categories is that C1 refers to teacher language and behaviors that might influence children's *becoming* curious by recognizing and preferring uncertainty, while C2 refers to teacher language and behaviors that might develop behaviors to resolve uncertainty, supporting students' information-seeking strategies and skills. To promote curious behavior, learning experiences can include opportunities to explore materials, ideas, and questions without concrete steps to follow, such as in a "tell and practice" approach, which research shows can "overemphasize efficiency at the expense of discovering new ways of seeing and doing" (Schwartz et al., 2012). This allows students to practice identifying the specific problem they are trying to solve and ways of solving it, with focus more on the process of exploring and learning over the end-goal of finding the answer (Dean and Kuhn, 2007; Sullivan, 2011). A second method of promoting curious behavior is to provide scaffolding to guide students' information-seeking, helping to make knowledge gaps less intimidating by breaking down steps to find information or even to help identify information needed and ideas for ways of getting it (Turner et al., 1998; van de Pol et al., 2010). Support for focusing on this learning process can be important for developing openness to uncertainty, as a strong focus on outcomes over the process is negatively associated with children's developing motivational framework (Park et al., 2016), while focusing on effort positively impacts children's motivational frameworks (Gunderson et al., 2013). Finally, teachers can simply provide positive reinforcement to students' curiosity and question asking, which can help create a classroom climate where curiosity is valued (Spargo and Enderstein, 1997; Pianta et al., 2008). Research shows that process praise (as opposed to praise for performance) is associated with children focusing on learning for the sake of learning over academic performance (Gunderson et al., 2018).

S1-2. Curiosity Suppression

While the absence of curiosity promotion may not have a negative impact on curiosity, we identified two instructional practices as potentially suppressive of students' curiosity. The first category of curiosity suppression includes instances whereby teachers model avoidance or discomfort with uncertainty or mistakes or make comments that suggest to students that mistakes should be avoided. Teachers' responses to student mistakes show how these responses can influence students' perceptions about errors (Käfer et al., 2019). For example, Tulis (2013) identified and measured teacher responses to students' mistakes as maladaptive or adaptive. The former, which included responses like criticism, negative affect (e.g., grimacing), or humiliation, made up about one third of teacher responses to mistakes, and there were rarely instances of reinforcing ideas around mistakes as learning opportunities. Not surprisingly, students tended to show negative affect when responses to mistakes, included humiliation, ridicule, or criticism (Tulis, 2013). On the other hand, students' perceptions of classroom climate as accepting of mistakes are associated with higher motivation (Käfer et al., 2019).

The second category of curiosity suppression includes teachers' discouragement of information-seeking behavior, such as negatively responding when students might try to do something in a different way than instructed or asking questions that went outside of the scope of the teachers' learning goals (Carlsen, 1991; Engel, 2015). Both students and teachers described barriers to question asking at school including questions being perceived as unwelcome interruptions or getting negative responses to questioning (Dogan and Yucel-Toy, 2021). Importantly, this category does not include teacher redirection to focus students' attention on instruction, as we see this practice as general classroom management to keep students from going off-task rather than curiosity suppression.

Stage 2. Developing the CiC Conceptual Framework Coding Protocol

Developing and Refining Initial Items

In developing items for coding observations based on the CiC framework, we used a similar iterative observation procedure with our team observing a standardized set of video-recorded classroom instruction from an unrelated research study. The team had weekly discussions while watching videos together about what types of language and behavior aligned with which framework practices, and how to disambiguate practices that were overlapping. For example, when a teacher asks students to think of alternative strategies or examples after an initial response is given, this could be considered an additional opportunity to think and participate, but is better fit as encouraging alternative ideas because it is more specific. We decided on a rule to code each instance as only one category, but to use the most specific category that fit what was observed. The specific examples of practices were recorded, along with detailed descriptions of the practices that were updated with discussions to create the CiC framework coding protocol.

Revising the Protocol to Be Comprehensive and Specific

After reaching agreement in the group observation coding sessions, the development team coded an additional set of five videos independently, and then met to discuss what was coding and further refine the protocol to include what was and what was not coded under different categories, and information about common questions and how coding decisions were made, all of which was included in the protocol document revisions. The independent coding followed by group discussions was an iterative process, continuing with about five additional videos in each session until the group discussions revealed agreement across coders. This resulted in a protocol that includes background literature and explanation of the purpose of the protocol, and an explanation of each framework component with detailed descriptions of the component, examples of what is coded as that component, and examples of what is not coded under that component. Additional notes were added based on the iterative discussions that revealed nuances to the coding protocol and helpful considerations for achieving reliable coding.

When coding, language excerpts were only included in one curiosity category. As discussed above, it was necessary in some cases to choose the *best* fit among the available categories, identified by alignment to and specificity of the category, with preference for higher specificity. In other instances, the same instructional activity might include more than one category if there are separate actions or components, for example, a teacher asks students to generate questions with a partner, and then has them share the questions as a whole group and write examples on the board [C1.3: prompting student questions; C1.1: opportunities to participate (pairs); C1.1: opportunities to participate (whole group)]. Likewise, instruction could be coded under more than one category if it contains multiple distinct components. For example, a teacher might say "Great question! Can anyone else think of more questions about this?" (this would be broken into two segments, coded as C2.3: positive response to a question and C1.3: prompting student questions). Note that not everything that a teacher says or does was coded; only instances that fit within the curiosity and motivation categories were coded. The full coding protocol is publicly available on databrary.org.[1]

Observer Training

Once agreement was reached by the development team and the protocol was finalized, a new team of data coders were trained to use the protocol to test the feasibility of using it for future research. Training coders to use the CIC was a multistep process. The first step involved having them read the protocol and test themselves on the specific categories using provided written practice activities and keys. Once coders were confident in understanding the individual categories and knew the differences across categories, they watched specific examples of classroom instruction using the protocol overview to practice recognizing instances of different categories, reviewing the protocol as needed. Coders then tested themselves using two training videos, involving coding videos using the protocol and then checking their codes against provided keys, which provide explanation for why different instructional components are coded (or not) as specific different categories. If they had mistakes or did not understand codes on the key, they completed an additional training video after more training.

Coding Methods and Efficiency

This initial work used consensus coding for reliability, in which multiple trained coders code the same videos independently, and then meet to ensure consensus and discuss any discrepant codes until 100% agreement was reached. We chose this method because the sample in the feasibility test was planned to be small, and we wanted to have the opportunity to evaluate and revise the protocol if needed. However, we designed the protocol with plans for using another more efficient method in future work, with individual coders trained to reliable coding standards. For example, the coders might code three test videos and have these scored by the protocol developers against a key, achieving a minimum of 85% accuracy of coding. Accuracy would be determined by the match between frequency across categories with the key. Then, reliability would be assessed across coders by having a subset of videos coded by multiple coders so that the consistency across coders could be assessed. Consensus coding is labor-intensive, so ideally reliability coding will be used in future work. **Table 1** describes the coding categories used with example instructional language for each coding category.

Testing the CiC Framework

Video Data Sources

Our data source for this initial test of the CiC framework used a convenience sample, the Measures of Effective Teaching (MET) database (Bill and Melinda Gates Foundation, 2009-2011), a corpus of classroom videos developed for researchers to study how instruction can be most effective for student outcomes. Because our focus of the coding was on instructional language, it was critical that videos have adequate audio data to clearly hear what is being said. Many constructs that were coded required some understanding of the general activity and/or context of the instruction and lesson, so having a view of materials, teacher gestures, projected material or whiteboard, etc. was beneficial to having a coherent understanding of the instruction for coding. For these reasons, we included videos rated as high-quality across all A/V quality ratings by MET research staff for teachers who consented for secondary data analyses (if teachers had multiple videos fitting all criteria,

[1]https://nyu.databrary.org/volume/1377.

we used the earliest one). Finally, we only included videos of instruction collected in year 2 of the MET study, so that we could use student survey items that were only collected in year 2. These filters resulted in 35 videos.

Goals for the Feasibility Testing and What Was Coded

Our goal in coding with the CiC protocol was to estimate the frequency and range of experiences that students have related to promoting and suppressing curiosity. We aimed to develop a protocol that could be used in a relatively short amount of time, but that would be representative of what students in a classroom *typically* experience. In the future, we expect this coding method to be useful for real-time coding in classrooms. For the current work, we coded 10-min time segments from each lesson, with many lessons including two segments that were averaged, including one from the beginning and one from the middle of the lesson.

Selecting and Preparing Observations for Coding

Because our initial goal was to test the framework and whether it could be used to reliably code for curiosity promotion, we focused on a single domain to limit content-specific variability. Most research has examined curiosity as a domain-general construct, but recent work proposes that curiosity should be examined in domain-specific contexts (Peterson and Cohen, 2019). Understanding domain-specific curiosity may better inform how curiosity leads to learning in domains, such as math or science. In science lessons, for instance, recognizing and responding to uncertainty can drive science learning (Manz and Suárez, 2018; Lamnina and Chase, 2019). In math, highlighting knowledge gaps and providing opportunities to explore may be particularly important (Peterson and Cohen, 2019; Rumack and Huinker, 2019). As the current framework examines several specific curiosity-promoting practices, it can potentially be used to examine both domain-general and -specific curiosity.

We chose math as the first domain to investigate and focused on upper-elementary grades for several reasons. First, interest in and attitudes about academic subjects and ability, including stereotype ideas about gender differences, develop very early. By early elementary grades, children already show stereotypical gender beliefs of a stronger association between boys and math than girls and math, with girls showing a weaker self-identification with math than boys (Cvencek et al., 2011). Thus, finding ways of improving math instruction and motivation in learning could provide beneficial implications for addressing the problem of encouraging girls to pursue math and math-intensive careers. Additionally, math interest declines for all students between late elementary and middle school, though girls begin at much lower interest levels than boys in 4-5th grade (62% of girls and 95% of boys giving a positive interest response; (LeGrand, 2013). Teachers' enthusiasm is associated with math interest even during the periods of math interest decline observed in adolescence, suggesting the important role of teacher practices (Frenzel et al., 2010). We were therefore interested in the role instruction might play in promoting children's curiosity at the critical point immediately preceding the transition from elementary to middle school (4–5th grade), when interests have been observed to decline. To control for the difference in opportunity to promote curiosity and motivation based on age and topic, we selected videos within a single grade level that were focused on a specific topic: adding and subtracting fractions. Because understanding teacher language was necessary for our coding protocol, we selected videos that were identified by the MET project as high-quality across all A/V quality ratings by MET research staff for teachers who consent for secondary data analysis (if teachers had multiple videos fitting all criteria, we used the earliest one), and then a research assistant screened videos to ensure adequate quality to hear and understand the teacher for two, 10-min segments.

Segments were selected to begin at points in which a teacher was beginning instruction on content—we did not begin a segment while the teacher was asking students to take out their homework or textbook, or if the teacher was reading the information about the video recorders, etc. We also ensured that the 10-min segments did not include any type of non-typical activity, for example, a test during which there would not be an expectation of instruction, and/or something that was not expected to occur more than once a week. These activities were labeled as uncodable time, and if they occurred, the 10-min segment was selected to begin after the uncodable time ended. A single coder prescreened all videos and identified the codable segments by time points to begin coding, which were then used by all coders. Because videos were selected based on having high ratings of both audio and video coding, most teacher language was audible and codeable.

To begin exploring the validity of the coding protocol, we looked at questions students asked in the coded segments to test if they were associated with frequency of curiosity-promoting practices. Student questions were transcribed from a subset of videos as possible (e.g., audio quality was sufficient for student voices), resulting in 57 total student questions asked across 45 coded segments ($N = 23$ lessons; one lesson only had one 10-min segment with sufficient audio).

FEASIBILITY TEST RESULTS: CURIOSITY PROMOTION

Two 10-min segments were successfully coded from each of 35 5th-grade math lessons on adding and subtracting fractions using the video screening method described above. In general, and as many would suspect, we observed extremely few instances of curiosity-promoting language. **Table 2** presents the total number of times that each curiosity-promoting type was observed across all videos, the mean frequency of observed instances within a lesson (two coded segments), the range of frequency observed across lessons, and the percentage of teachers who had at least one occurrence observed. While most teachers were observed using one or more instances of promoting recognition of or comfort with uncertainty (83%), the frequency of engaging in this across the two coded segments of a lesson was only 2.03 instances ($SD = 2.1$). Instruction that included

TABLE 2 | Observations of curiosity-promoting instruction across all lessons coded (two 10-min segments from each of 35 lessons).

Coding Category:	Total (all) observation	Mean frequency per segment	Range of frequencies	% of teachers (any coded)
C1.1 Provide opportunities to think, question, participate	45	1.29 (SD = 1.89)	0–10	60%
C1.2 Modeling own comfort with uncertainty	3	0.09 (SD = 0.28)	0–1	9%
C1.3 Prompting question Generation	0	0	0	0%
C1.4 Reflecting on student prior knowledge and uncertainty	8	0.23 (SD = 0.49)	0–2	20%
C1.5 Encouraging alternative Ideas	15	0.43 (SD = 0.92)	0–5	31%
C1: Average of promoting comfort with uncertainty		2.03 (SD = 2.13)	0–11	
C2.1 Provide opportunities to explore and "figure out"	3	0.09 (SD = 0.37)	0–2	6%
C2.2 Scaffolding and guidance in resolving uncertainty	1	0.03 (SD = 0.17)	0–1	3%
C2.3 Positive responses to questions asked	7	0.20 (SD = 0.47)	0–2	17%
C2: Average of supporting exploration and questioning		0.31 (SD = 0.63)	0–2	
S1: Promoting discomfort with uncertainty	1	0.03 (SD = 0.17)	0–1	3%
S2: Negative responses to curiosity and information-seeking	7	0.20 (SD = 0.87)	0–5	9%

opportunities to be curious, such as opportunities to "figure out" or positive reinforcement of curious behavior was only observed in 23% of teachers, with a mean of only 0.3 (*SD* = 0.63) instances total across the lessons observed. Most striking, we did not observe a single instance of teachers prompting students to generate questions across all videos coded. Note that a teacher simply checking understanding by asking, "any questions?" was not coded; rather this code was for explicitly prompting students to generate questions (e.g., "what questions do you have?" or "can you think of any questions we could ask?"). These low numbers indicate that promoting curiosity promotion was infrequent, but curiosity suppression was even rarer, with only 9% of teachers having any instances, and occurring 0.23 (*SD* = 0.91) times per lesson observed, meaning we only saw 7 instances across three of the 35 observations.

As an initial exploration of effects of curiosity promotion on students, we tallied questions asked by students during the coded segments. Surprisingly, we did not observe a relation between curiosity-promoting language and student questions in class. When we explored further, we found that when students heard no curiosity-promoting language, they asked an average of 1.4 (*SE* = 0.12) questions; when they heard only one or two instances of curiosity-promoting language, they only asked 0.4 questions (*SE* = 0.11); however, when they heard more than two instances of curiosity-promoting language, they asked an average of 2.1 (*SE* = 0.14) questions. It is important to note that this is at the classroom level—two questions means two questions asked *by the entire class*, in the whole 10-min segment. In fact, of the 45 segments for which we were able to transcribe student questions, more than half (24 total) had no student questions asked. Analyzed at the class-level, these differences are not significant. However, we only counted the frequency of questions and did not code questions by type, and coders reported that many questions were simply asking permission or clarification, so may not have been indicative of student curiosity.

DISCUSSION

Prior research shows a need to promote students' curiosity, and for explicit attention to practices that create a curiosity-supportive classroom climate, such as concrete support for and encouragement of curiosity in students (Post and Walma van der Molen, 2018). Classroom climate is impacted by students' observations of what teachers care about, which can be portrayed both indirectly and directly, such as through instruction (Jirout et al., 2018). As Katz notes, "...Even very young children are most likely making inferences about what adults care about based on multiple observations of the adults' actual behavior in context" (Katz, 1998, p. 38). This paper described our process of developing a framework and observational tool to study how specific instructional practices can promote students' curiosity by supporting their comfort with and recognition of things they do not know, which can help to promote their becoming curious, and by developing their information-seeking skills which can promote their curious behavior and learning. Although many studies have focused on individual aspects of instruction that might independently support curiosity, much of this research occurred in lab-based research and involved manipulating instruction. The CiC framework coding protocol will allow for understanding what is happening in actual classrooms and the study of how instructional practices impact student curiosity in these classrooms.

This framework and protocol are not intended to reveal what a "good" teacher or "good" teaching looks like; rather, our goal was to understand what instructional practices could occur that support developing student curiosity specifically. We tested the feasibility of using the developed tool for observing for curiosity promotion and described how frequently instructional practices related to promotion and suppression of curiosity occur. In addition to the protocol itself, we described this process of developing it and the results of the feasibility test. We discuss each of these components below, followed by suggestions for future research to validate and use the developed tool.

CiC Design and Development Process

There are not currently well-developed methods of assessing what teachers do (or do not do) to support student curiosity, which limits the ability of research to support educational promotion of curiosity and research on the value of doing this. The work described here takes the initial step toward understanding the influence on instruction on student curiosity by developing a method for measuring instructional practices that might promote student curiosity to assess what works and whether it matters for motivation and learning. This work began with a process of integrating prior theory and research to align with our operationalization of curiosity to understand what changes in curiosity in a classroom context would look like, which included identifying the two types of support we include: support for becoming curious (i.e., identifying things one does not know and wants to find out) and support for being curious (i.e., seeking information). We then used an iterative process of observations, discussions, and alignment to prior work to identify the items in the CiC framework using instruction identified as high-quality, followed by a similar process using observations more representative of typical educational contexts to explain in detail what the framework practices could look like in observations of classroom instruction and methods for training observers and conducting these observations.

The process of developing this framework and protocol was difficult, and we hope that in addition to the contribution of the CiC itself, that the description of this process will also be helpful for researchers. Specifically, we would summarize our experience in three lessons learned for future efforts in similar work. First, operationalizing the construct of interest and being specific in what we wanted to observe was important. Although curiosity seems like a simple and common construct, it is poorly understood with definitional and measurement challenges (Jirout and Klahr, 2012; Kidd and Hayden, 2015). Despite these challenges, we used a specific operationalization of curiosity linked to learning behaviors to consider how student curiosity can be influenced by instructional practices. This operationalization focused our efforts around identifying practices related to developing comfort for and recognition of uncertainty and information-seeking, which helped to identify and categorize themes among the instruction practices that we observed in developing the framework. A second lesson was that developing this protocol was a much more iterative process than we originally anticipated. Instruction varies greatly from teacher to teacher and classroom to classroom, and it took many revisions of our framework descriptions and protocol instructions to become a generalization of instructional practices that aligned to the framework while still reflecting the many different types of instruction that might fit each practice. Our third lesson was related to the second: this was not a quick development process that could be assumed to be easily completed at the beginning of a larger project. This process was extensive, lasting about 18 months, because of the need for countless discussions, iterations, and revisions. Had we been pressured to complete the process of developing the protocol to conduct research using it, we may not have been able to spend the needed time and effort in its development.

Results of the Feasibility Test

We acknowledge that the work here is presented as a very initial step in understanding the role of curiosity in natural settings, in that our current work aimed only to test the protocol and describe what types and frequency of curiosity promotion was observed in a single domain and grade level. Future work to validate the tool is necessary. That said, observations using the developed protocol showed low levels of curiosity-promoting instruction in a small feasibility test, despite prior research showing broad agreement that curiosity is valued and important in learning contexts. The practices observed were focused on ways that teachers promoted comfort with and desire to explore uncertainty. Children need opportunities to become curious and practice being curious, and this can happen by supporting and promoting students' comfort with uncertainty (Jirout et al., 2018). If students are expected to listen and learn information without pauses to think about the information and ideas, they would not have the time needed for reflecting beyond what it is they heard to consider what it is they do not know but could be curious about, what Glăveanu (2022) refers to as "uncertain knowing." Further, supporting a mindful approach to thinking about uncertainty can help to open children's thinking and reduce the focus on worrying about judgment (Henriksen et al., 2022). The observational protocol described here, based on the CiC framework, will allow future research to explore what kinds of curiosity-promoting instructional practices are happening, how frequently, and whether those practices are linked to student outcomes. Further, the framework provides specific instructional practices that can be individually tested and explored in controlled studies for their efficacy in influencing students' curiosity to develop concrete practical implications for what educators can do to support students' curiosity, such as using "structured uncertainty" to support curiosity (Beghetto, 2020), where scaffolding and support is built in for students to practice thinking in new or different ways about a problem.

Importance of and Need for Research on Curiosity Promotion

The findings of Post and colleagues (2018) that students do not see being curious as consistent with learning in school is troubling, but it offers a clear need for future research to understand how we can change this perception. Promoting students' curiosity to learn in educational contexts could make learning more enjoyable and support motivation (Midgley et al., 2001), which could support future learning (Jirout et al., in review), and it also could positively influence learning behaviors, such as question asking and class participation (D. Park et al., 2017; Jirout and Klahr, 2020). This might be especially important and effective for domains like science (Jirout, 2020), where children's ability to ask and think about questions is seen as fundamental, with the inclusion of question asking as the first of eight scientific and engineering practices that span across all grade levels and content areas in the National Research Council's National Science Education Standards (NRC, 2012).

Although the current test only looked at math instruction, we were still surprised not to have observed a single instance of teachers prompting students to generate questions, which could be an important way to help them recognize things they might be curious to know. In exploring their own curiosity, students might also develop more sustained interests in topics (Hidi and Renninger, 2006), which can support learning more complex material and contribute to learning beyond curiosity (Hidi and Renninger, 2019). In student self-reports of their interest across domains, science has the lowest proportion of being considered most interesting compared to math and reading, and this is significantly lower for females than males and for students eligible for free and reduced school lunch compared to higher-income peers (U.S. Department of Education, 2015). Importantly, these students' ratings of interest in science are significantly associated with their performance (U.S. Department of Education, 2015). Developing curiosity could help to promote learning across domains and could especially support retaining and motivating students in pursuing science careers.

Limitations and Future Directions

Although we observed low levels of these practices, we looked only within a narrow scope, assessing a small sample of lessons within the single domain of math, so we caution readers in drawing too much meaning from the data presented, which was intended more as a test of feasibility of using the protocol than a description of what is currently happening in classrooms. Future work should look more broadly at different educational levels and across different academic subjects and types of instruction (e.g., whole class and small group). Just as there is a need to explore whether students' curiosity is domain-general (e.g., Loewenstein, 1994; Markey and Loewenstien, 2014), domain-specific (Peterson and Cohen, 2019), or, as we expect, reflects evidence of both domain-general and domain-specific curiosity (Murayama et al., 2019), it is possible that curiosity promotion might look different across domains or even across education levels.

We also only provide descriptive information about what we observed. While we attempted to look at associations with student questioning, there were too few questions asked by students across the observed lessons. Future research should explore what outcomes are predicted by curiosity-promoting instructional practices, especially looking at changes in students' curiosity and related constructs, which was not possible using the convenience sample included here. Additionally, there are likely other important factors to consider in this future work. Though many metacognitive and social–emotional factors are important for children's successful learning, this paper focused on curiosity as one such factor in that it can both move one to act and direct behavior toward finding specific information (Wentworth and Witryol, 2003). Curiosity can help to develop sustained interests (Hidi and Renninger, 2020), and, in turn, promote self-regulation, information-seeking, and motivation (Renninger, 2000; Hidi and Renninger, 2006). There are likely many other additional factors that are also important for curiosity and learning, such as mindset, achievement goal orientation, self-efficacy, academic courage, and intellectual humility. Understanding how curiosity relates to and differs from these different constructs is another important area for future work. In addition, there are likely important aspects of the social context of classrooms and peer interactions that also can influence curiosity (Käfer et al., 2019), as well as classroom resources and instructional design factors that may differ, such as use of and familiarity with technology. We hope that this work advancing methods of studying curiosity promotion in instruction will be useful in future research to explore these factors and the associations among them to understand how educational contexts can support students' curiosity. The CiC framework and coding protocol, as well as tools, such as observation sheets, are openly available and can be used to pursue many of these and other future directions.

Conclusion

This work describes the development of an observational method to assess frequency of curiosity promotion in classroom instruction. The logical next step is to use it to assess how often students experience curiosity-promoting instruction and the ways in which these experiences foster or suppress motivation, learning, engagement, and achievement. Prior work has suggested that student curiosity during classroom learning is low and decreases with grade level (Engel, 2011), and the CiC Framework coding protocol will allow researchers to observe whether curiosity promotion is happening in classrooms and whether or not it has an impact on student curiosity and learning. We describe a feasibility test of the observational protocol using video-recorded classroom instruction from a nationally representative database, showing that it is possible to measure these instructional practices. While the levels we observed in the feasibility test were low, further research is needed to explore other contexts, as well as whether this varies by subject, teacher, grade level, and many other factors. The protocol described provides a methodological tool to advance this important and needed research.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Virginia Institutional Review Board - Social and Behavioral Sciences. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JJ wrote the grant proposals with the initial plan and idea for this work. JJ, SZ, and VV developed the framework and led the observational protocol work described here. JJ and SZ wrote the manuscript with important contributions from NE and VV. All authors contributed to the article and approved the submitted version.

FUNDING

The research described here was supported by grants from the Spencer Foundation and the Center for Curriculum Redesign to JJ. This article was also made possible through grants from the Templeton Foundation and the Jacobs Foundation supporting this team's broader program of research on curiosity promotion. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agencies.

ACKNOWLEDGMENTS

The authors are grateful to Judy Paulick, Marissa Drell, Karime Cameron, and Jenny Lee for their contributions to the work described here.

REFERENCES

Beghetto, R. A. (2020). "Uncertainty," in *The Palgrave Encyclopedia of the Possible*. ed. V. Glăveanu (Cham: Palgrave Macmillan), 1–7.

Bill and Melinda Gates Foundation. (2009-2011). Measures of Effective Teaching: 1 - Study Information. Inter-university Consortium for Political and Social Research [distributor], 2018-09-19. doi: 10.3886/ICPSR34771.v3

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., and Schulz, L. (2011). The double-edged sword of pedagogy: instruction limits spontaneous exploration and discovery. *Cognition* 120, 322–330. doi: 10.1016/j.cognition.2010.10.001

Brophy, J. E., and Good, T. (1984). *Teacher Behavior and Student Achievement No. 73*. East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Carlsen, W. S. (1991). Subject-matter knowledge and science teaching: a pragmatic perspective. *Adv. Res. Teach.* 2, 115–143.

Chak, A. (2007). Teachers' and parents' conceptions of children's curiosity and exploration. *Int. J. Early Years Educ.* 15, 141–159. doi: 10.1080/09669760701288690

Ciardiello, A. V. (1998). Did you ask a good question today? Alternative cognitive and metacognitive strategies. *J. Adolesc. Adult. Lit.* 42, 210–219.

Coie, J. D. (1974). An evaluation of the cross-situational stability of children's curiosity. *J. Pers.* 42, 93–116. doi: 10.1111/j.1467-6494.1974.tb00559.x

Cook, C., Goodman, N. D., and Schulz, L. E. (2011). Where science starts: spontaneous experiments in preschoolers' exploratory play. *Cognition* 120, 341–349. doi: 10.1016/j.cognition.2011.03.003

Cunnigham, C. M. (1997). Who knows?: The influence of teachers' sociological understanding of science (SUS) on knowledge, authority, and control in the classroom. *J. Classroom Interact.* 32, 24–34.

Cvencek, D., Meltzoff, A. N., and Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Dev.* 82, 766–779. doi: 10.1111/j.1467-8624.2010.01529.x

Danovitch, J. H., Mills, C. M., Sands, K. R., and Williams, A. J. (2021). Mind the gap: how incomplete explanations influence children's interest and learning behaviors. *Cogn. Psychol.* 130:101421. doi: 10.1016/j.cogpsych.2021.101421

Dean, D. Jr., and Kuhn, D. (2007). Direct instruction vs. discovery: the long view. *Sci. Educ.* 91, 384–397. doi: 10.1002/sce.20194

Dess, G., and Picken, J. (2012). Changing roles: leadership in the 21st century. *Organ. Dyn.* 28, 18–34. doi: 10.1016/S0090-2616(00)88447-8

Dogan, F., and Yucel-Toy, B. (2021). Students' question asking process: a model based on the perceptions of elementary school students and teachers. *Asia Pac. J. Educ.*, 1–16. doi: 10.1080/02188791.2021.1873104

Duschl, R. A., and Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Stud. Sci. Educ.* 38, 39–72. doi: 10.1080/03057260208560187

Engel, S. (2011). Children's need to know: curiosity in schools. *Harv. Educ. Rev.* 81, 625–645. doi: 10.17763/haer.81.4.h054131316473115

Engel, S. (2015). *The Hungry Mind: The Origins of Curiosity in Childhood*. Cambridge, MA: Harvard University Press.

Engel, S., and Randall, K. (2009). How teachers respond to children's inquiry. *Am. Educ. Res. J.* 46, 183–202. doi: 10.3102/0002831208323274

Engelhard, G., and Monsaas, J. A. (1988). Grade level, gender, and school-related curiosity in urban elementary schools. *J. Educ. Res.* 82, 22–26. doi: 10.1080/00220671.1988.10885860

Fredrickson, B. L. (1998). What good are positive emotions? *Rev Gen Psychol.* 2, 300–319. doi: 10.1037/1089-2680.2.3.300

Fredrickson, B. L., and Joiner, T. (2002). Positive emotions trigger upward spirals toward emotional well-being. *Psychol. Sci.* 13, 172–175. doi: 10.1111/1467-9280.00431

Frenzel, A. C., Goetz, T., Pekrun, R., and Watt, H. M. (2010). Development of mathematics interest in adolescence: influences of gender, family, and school context. *J. Res. Adolesc.* 20, 507–537. doi: 10.1111/j.1532-7795.2010.00645.x

Glăveanu, V. P. (2022). "Not knowing," in *Uncertainty: A Catalyst for Creativity, Learning and Development*. eds. R. A. Beghetto and J. G. Jaeger Berlin: Springer International Publisher.

Gordon, G., Breazeal, C., and Engel, S. (2015). Can children catch curiosity from a social robot? in Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction. 91–98.

Grolnick, W. S., and Ryan, R. M. (1987). Autonomy in children's learning: an experimental and individual difference investigation. *J. Pers. Soc. Psychol.* 52, 890–898. doi: 10.1037/0022-3514.52.5.890

Gunderson, E. A., Donnellan, M. B., Robins, R. W., and Trzesniewski, K. H. (2018). The specificity of parenting effects: differential relations of parent praise and criticism to children's theories of intelligence and learning goals. *J. Exp. Child Psychol.* 173, 116–135. doi: 10.1016/j.jecp.2018.03.015

Gunderson, E. A., Gripshover, S. J., Romero, C., Dweck, C. S., Goldin-Meadow, S., and Levine, S. C. (2013). Parent praise to 1- to 3-year-olds predicts children's motivational frameworks 5 years later. *Child Dev.* 84, 1526–1541. doi: 10.1111/cdev.12064

Gweon, H., Pelton, H., Konopka, J. A., and Schulz, L. E. (2014). Sins of omission: children selectively explore when teachers are under-informative. *Cognition* 132, 335–341. doi: 10.1016/j.cognition.2014.04.013

Hamre, B. K., and Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Dev.* 76, 949–967. doi: 10.1111/j.1467-8624.2005.00889.x

Henriksen, D., Richardson, Gruber, N., and Mishra, P. (2022). "The uncertainty of creativity: opening possibilities and reducing restrictions through mindfulness," in *Uncertainty: A Catalyst for Creativity, Learning and Development*. eds. R. A. Beghetto and J. G. Jaeger Berlin: Springer International Publisher.

Hidi, S., and Renninger, K. A. (2006). The four-phase model of interest development. *Educ. Psychol.* 41, 111–127. doi: 10.1207/s15326985ep4102_4

Hidi, S. E., and Renninger, K. A. (2019). Interest development and its relation to curiosity: needed neuroscientific research. *Educ. Psychol. Rev.* 31, 833–852. doi: 10.1007/s10648-019-09491-3

Hidi, S. E., and Renninger, K. A. (2020). On educating, curiosity, and interest development. *Curr. Opin. Behav. Sci.* 35, 99–103. doi: 10.1016/j.cobeha.2020.08.002

Jean, A., Daubert, E., Yu, Y., Shafto, P., and Bonawitz, E. (2019). "Pedagogical questions empower exploration," in *Proceedings of the Annual Conference of the Cognitive Science Society*. Montreal, Canada.

Jirout, J. J. (2020). Supporting early scientific thinking through curiosity. *Front. Psychol.* 11:1717. doi: 10.3389/fpsyg.2020.01717

Jirout, J., and Klahr, D. (2012). Children's scientific curiosity: in search of an operational definition of an elusive concept. *Dev. Rev.* 32, 125–160. doi: 10.1016/j.dr.2012.04.002

Jirout, J., and Klahr, D. (2020). Questions – and some answers – about young children's questions. *J. Cogn. Dev.* 21, 729–753. doi: 10.1080/15248372.2020.1832492

Jirout, J., Ruzek, E., Vitiello, V., Whittaker, J. V., and Pianta, R. (Revision under review). School enjoyment and the relation to the development of general knowledge.

Jirout, J., Vitiello, V., and Zumbrunn, S. (2018). "Curiosity in Schools," in *The New Science of Curiosity*. ed. G. Gordon (Hauppauge, NY: Nova Science Publishers).

Käfer, J., Kuger, S., Klieme, E., and Kunter, M. (2019). The significance of dealing with mistakes for student achievement and motivation: results of doubly latent multilevel analyses. *Eur. J. Psychol. Educ.* 34, 731–753. doi: 10.1007/s10212-018-0408-7

Kashdan, T. B. (2006). Exploring the functions, correlates, and consequences of interest and curiosity. *J. Pers. Assess.* 87, 352–353. doi: 10.1207/s15327752jpa8703_15

Kashdan, T. B., and Steger, M. F. (2007). Curiosity and pathways to well-being and meaning in life: traits, states, and everyday behaviors. *Motiv. Emot.* 31, 159–173. doi: 10.1007/s11031-007-9068-7

Katz, L. G. (1998). "What can we learn from Reggio Emilia?" in *The Hundred Languages of Children*. eds. C. Edwards, L. Gandini and G. Forman (Westport, CT: Greenwood Publishing Group), 27–45.

Keller, H., Schneider, K., and Henderson, B. (eds.). (2012). *Curiosity and Exploration*. Berlin: Springer Science & Business Media.

Kidd, C., and Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron* 88, 449–460. doi: 10.1016/j.neuron.2015.09.010

Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2012). The goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One* 7:e36399. doi: 10.1371/journal.pone.0036399

King, A. (1994). Guiding knowledge construction in the classroom: effects of teaching children how to question and how to explain. *Am. Educ. Res. J.* 31, 338–368. doi: 10.3102/00028312031002338

Lamnina, M., and Chase, C. C. (2019). Developing a thirst for knowledge: how uncertainty in the classroom influences curiosity, affect, learning, and transfer. *Contemp. Educ. Psychol.* 59:101785. doi: 10.1016/j.cedpsych.2019.101785

LeGrand, J. (2013). Exploring gender differences across elementary, middle, and high school students' science and math attitudes and interest [Ed.D., Northeastern University]. Available at: https://www.proquest.com/docview/1330500629/abstract/30316FD172924914PQ/1 (Accessed February 13, 2022)

Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personal. Individ. Differ.* 44, 1585–1595. doi: 10.1016/j.paid.2008.01.014

Litman, J. A., Collins, R. P., and Spielberger, C. D. (2005). The nature and measurement of sensory curiosity. *Personal. Individ. Differ.* 39, 1123–1133. doi: 10.1016/j.paid.2005.05.001

Loewenstein, G. (1994). The psychology of curiosity: a review and reinterpretation. *Psychol. Bull.* 116, 75–98. doi: 10.1037/0033-2909.116.1.75

Manz, E., and Suárez, E. (2018). Supporting teachers to negotiate uncertainty for science, students, and teaching. *Sci Ed.* 102, 771–795. doi: 10.1002/sce.21343

Markey, A., and Loewenstein, G. (2014). "Curiosity," in *International handbook of emotions in education*. Routledge. 238–255.

Martin, A. (2011). Courage in the classroom: exploring a new framework predicting academic performance and engagement. *Sch. Psychol. Q.* 26, 145–160. doi: 10.1037/a0023020

Martin, A. J., and Marsh, H. W. (2003). Fear of failure: friend or foe? *Aust. Psychol.* 38, 31–38. doi: 10.1080/00050060310001706997

Matuk, C., and Linn, M. C. (2015). "Examining the real and perceived impacts of a public idea repository on literacy and science inquiry." in *CSCL'15: Proceedings of the 11th International Conference for Computer Supported Collaborative Learning Vol. 1*, 150–157. Gothenburg, Sweden: International Society of the Learning Sciences.

McDonald, J. P. (1992). *Teaching: Making Sense of an Uncertain Craft*. New York: Teachers College Press.

McTighe, J., and Lyman, F. T. (1988). Cueing thinking in the classroom: the promise of theory-embedded tools. *Educ. Leadersh.* 45, 18–24.

Menning, S. F. (2019). Why nurturing curiosity is an ethical endeavour: exploring practitioners' reflections on the importance of curiosity. *Int. J. Early Years Educ.* 27, 34–51. doi: 10.1080/09669760.2018.1547632

Metcalfe, J., Kennedy-Pyers, T., and Vuorre, M. (2021). Curiosity and the desire for agency: wait, wait… don't tell me! *Cognit. Res. Principles Implications* 6, 1–8. doi: 10.1186/s41235-021-00330-0

Midgley, C., Kaplan, A., and Middleton, M. (2001). Performance-approach goals: good for what, for whom, under what circumstances, and at what cost? *J. Educ. Psychol.* 93, 77–86. doi: 10.1037/0022-0663.93.1.77

Murayama, K. (2022). A reward-learning framework of knowledge acquisition: An integrated account of curiosity, interest, and intrinsic–extrinsic rewards. *Psychol. Rev.* 129, 175–198. doi: 10.1037/rev0000349

Murayama, K., FitzGibbon, L., and Sakaki, M. (2019). Process account of curiosity and interest: A reward-learning perspective. *Educ Psychol Rev.* 31, 875–895. doi: 10.1007/s10648-019-09499-9

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: National Academies Press.

Park, D., Gunderson, E. A., Tsukayama, E., Levine, S. C., and Beilock, S. L. (2016). Young children's motivational frameworks and math achievement: Relation to teacher-reported instructional practices, but not teacher theory of intelligence. *J. Educ. Psychol.* 108:300.

Park, H. W., Rosenberg-Kima, R., Rosenberg, M., Gordon, G., and Breazeal, C. (2017). "Growing growth mindset with a social robot peer." in *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, 137–145. Vienna, Austria.

Park, D., Tsukayama, E., Goodwin, G. P., Patrick, S., and Duckworth, A. L. (2017). A tripartite taxonomy of character: evidence for intrapersonal, interpersonal, and intellectual competencies in children. *Contemp. Educ. Psychol.* 48, 16–27. doi: 10.1016/j.cedpsych.2016.08.001

Pekrun, R., and Linnenbrink-Garcia, L. (2012). "Academic emotions and student engagement," in *Handbook of Research on Student Engagement* eds. S. Christenson, A. Reschly and C. Wylie (Boston, MA: Springer), 259–282.

Peterson, E. G., and Cohen, J. (2019). A case for domain-specific curiosity in mathematics. *Educ. Psychol. Rev.* 31, 807–832. doi: 10.1007/s10648-019-09501-4

Pianta, R. C., La Paro, K. M., and Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. (pp. xi, 112). Baltimore, MD: Paul H Brookes Publishing.

Post, T., and Walma van der Molen, J. H. (2018). Do children express curiosity at school? Exploring children's experiences of curiosity inside and outside the school context. *Learn. Cult. Soc. Interact.* 18, 60–71. doi: 10.1016/j.lcsi.2018.03.005

Renninger, K. A. (2000). "Individual interest and its implications for understanding intrinsic motivation," in *Intrinsic and Extrinsic Motivation* (Cambridge, MA: Academic Press), 373–404.

Rosenshine, B., Meister, C., and Chapman, S. (1996). Teaching students to generate questions: a review of the intervention studies. *Rev. Educ. Res.* 66, 181–221. doi: 10.3102/00346543066002181

Rowe, M. B. (1972). "Wait-Time and Rewards as Instructional Variables: Their Influence on Language, Logic, and Fate." in *Presented at the National Association for Research in Science Teaching*, Chicago, Illinois, April 1972.

Rowe, M. B. (1986). Wait time: slowing down may be a way of speeding up! *J. Teach. Educ.* 37, 43–50. doi: 10.1177/002248718603700110

Rumack, A. M., and Huinker, D. (2019). Capturing mathematical curiosity with notice and wonder. *Math. Teach. Middle School* 24, 394–399. doi: 10.5951/mathteacmiddscho.24.7.0394

Ryan, R. M. (2012). *The Oxford Handbook of Human Motivation*. Oxford: Oxford University Press.

Schwartz, D. L., Chase, C. C., and Bransford, J. D. (2012). Resisting overzealous transfer: coordinating previously successful routines with needs for new learning. *Educ. Psychol.* 47, 204–214. doi: 10.1080/00461520.2012.696317

Shah, P. E., Weeks, H. M., Richards, B., and Kaciroti, N. (2018). Early childhood curiosity and kindergarten reading and math academic achievement. *Pediatr. Res.* 84, 380–386. doi: 10.1038/s41390-018-0039-3

Shneidman, L., Gweon, H., Schulz, L. E., and Woodward, A. L. (2016). Learning from others and spontaneous exploration: a cross-cultural investigation. *Child Dev.* 87, 723–735. doi: 10.1111/cdev.12502

Spargo, P. E., and Enderstein, L. G. (1997). What questions do they ask? *Sci. Child.* 34, 43–45.

Spektor-Levy, O., Baruch, Y. K., and Mevarech, Z. (2013). Science and scientific curiosity in pre-school—The teacher's point of view. *Int. J. Sci. Educ.* 35, 2226–2253. doi: 10.1080/09500693.2011.631608

Sullivan, F. R. (2011). Serious and playful inquiry: epistemological aspects of collaborative creativity. *J. Educ. Technol. Soc.* 14, 55–65.

Tizard, B., and Hughes, M. (1984). *Children Learning at Home and in School.* London: Fontana.

Tizard, B., Hughes, M., Carmichael, H., and Pinkerton, G. (1983). Children's questions and adults answers. *J. Child Psychol. Psychiatry* 24, 269–281. doi: 10.1111/j.1469-7610.1983.tb00575.x

Tulis, M. (2013). Error management behavior in classrooms: teachers' responses to student mistakes. *Teach. Teach. Educ.* 33, 56–68. doi: 10.1016/j.tate.2013.02.003

Turner, J. C., Meyer, D. K., Cox, K. E., Logan, C., DiCintio, M., and Thomas, C. T. (1998). Creating contexts for involvement in mathematics. *J. Educ. Psychol.* 90, 730–745. doi: 10.1037/0022-0663.90.4.730

U.S. Department of Education (2015). Students views of mathematics, reading, and science. National Assessment of Educational Progress (NAEP), various years, 2009–2015 Mathematics Assessments, 2002–2015 Reading Assessments, and 2009–2015 Science Assessments. Washington, DC: National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Dept. of Education.

van de Pol, J., Volman, M., and Beishuizen, J. (2010). Scaffolding in teacher–student interaction: a decade of Research. *Educ. Psychol. Rev.* 22, 271–296. doi: 10.1007/s10648-010-9127-6

van Schijndel, T. J. P., Jansen, B. R. J., and Raijmakers, M. E. J. (2018). Do individual differences in children's curiosity relate to their inquiry-based learning? *Int. J. Sci. Educ.* 40, 996–1015. doi: 10.1080/09500693.2018.1460772

von Stumm, S., Hell, B., and Chamorro-Premuzic, T. (2011). The hungry mind: intellectual curiosity is the third pillar of academic performance. *Perspect. Psychol. Sci.* 6, 574–588. doi: 10.1177/1745691611421204

Wentworth, N., and Witryol, S. L. (2003). "Curiosity, exploration, and novelty-seeking," in *Well-Being: Positive Development Across the Life Course*. eds. M. H. Bornstein, L. Davidson, C. L. M. Keyes and K. A. Moore (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 281–294.

Willard, A. K., Busch, J. T., Cullum, K. A., Letourneau, S. M., Sobel, D. M., Callanan, M., et al. (2019). Explain this, explore that: a study of parent–child interaction in a children's museum. *Child Dev.* 90, e598–e617. doi: 10.1111/cdev.13232

Yeager, D. S., Paunesku, D., Walton, G. M., and Dweck, C. S. (2013). "How can we instill productive mindsets at scale? A review of the evidence and an initial R&D agenda." in *White Paper Prepared for White House Meeting Excellence in Education: The Importance of Academic Mindsets*.

Yu, Y., Landrum, A. R., Bonawitz, E., and Shafto, P. (2018). Questioning supports effective transmission of knowledge and increased exploratory learning in pre-kindergarten children. *Dev. Sci.* 21:e12696. doi: 10.1111/desc.12696

Zimmerman, B. J., and Pike, E. O. (1972). Effects of modeling and reinforcement on the acquisition and generalization of question-asking behavior. *Child Dev.* 43, 892–907. doi: 10.2307/1127640

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

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

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 16 May 2022
doi: 10.3389/fpsyg.2022.841160

The Effect of Physical Exercise on Depression in College Students: The Chain Mediating Role of Self-Concept and Social Support

Junliang Zhang, Shuang Zheng and Zhongzheng Hu*

College of Science and Technology, Nanchang Hangkong University, Jiujiang, China

Objective: This study introduced self-concept and social support as research variables to establish a research mechanism, in order to encourage college students to participate in sports better, relieve or overcome depression.

Methods: The survey was conducted among 1,200 college students in Jiangxi, China. Serial mediation models were used to examine whether self-concept and social support mediated in the effect of physical exercise on depression.

Results: Physical exercise significantly negatively predicted college depression. Moreover, Self-concept and social support mediate the relationship between physical exercise and depression in college students.

Conclusion: This study reveals how physical exercise affects college students' depression and its mechanism, and the results have certain enlightenment significance for maintaining and promoting college students' physical and mental health.

Keywords: physical exercise, depression, self-concept, social support, college students

Edited by:
Antonio P. Gutierrez de Blume, Georgia Southern University, United States

Reviewed by:
Kwok-fai So, Jinan University, China
Juan De Dios Benítez Sillero, University of Córdoba, Spain

***Correspondence:**
Junliang Zhang
986012469@qq.com

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

Received: *22 December 2021*
Accepted: *21 March 2022*
Published: *16 May 2022*

Citation:
Zhang J, Zheng S and Hu Z (2022) The Effect of Physical Exercise on Depression in College Students: The Chain Mediating Role of Self-Concept and Social Support. Front. Psychol. 13:841160. doi: 10.3389/fpsyg.2022.841160

INTRODUCTION

College students are the backbone of social and economic development in the future, and their growth and mental health have been the focus of public attention.With the quickening pace of society, pressure from all aspects increases, if there is no good anti-pressure, they will lead to depression and other psychological problems. Depression is a mental state with reduced energy, low mood, loss of interest, and poor quality of life. In severe cases, suicidal ideation and even suicidal behavior may occur (Ibrahim et al., 2013). The depression discussed in this study is not the diagnostic depression of mental disease, and only refers to a sub-health state. Studies at home and abroad have found that the number of suicides due to depression has been increasing year after year in recent years (Walker et al., 2015). If the administrators in the university department fail to recognize this phenomenon and take appropriate intervention measures, intervention measures in time, the development of the physical and mental health of college students will be seriously hindered.

Physical exercise refers to any physical activity that promotes physical and mental development by means of physical movement (Morgan and Ellickson, 1989). People who engage in regular physical exercise have a high level of self-confidence and values, leading to greater satisfaction with life, and thus reduced anxiety and depression (Esmaeilzadeh, 2015; Wu et al., 2016). Rebecca et al. 's study on 467 adolescents showed that although physical activity had an intervention effect on adolescents' sleep quality and depression, it had no predictive effect on depression level (Slykerman et al., 2019). Salmon's study found that aerobic exercise has anti-depression and anti-anxiety effects, but strenuous physical exercise has a confusing effect on emotions, which has both positive and negative effects (Salmon, 2001). Strong evidence shows that habitual physical activity has not been proven to prevent depression, but increasing moderate physical activity can significantly reduce symptoms of depression and anxiety, and improve self-esteem and a positive outlook on life (Crone, 2003; Sun et al., 2014). Khanzada conducted a descriptive statistical analysis of 269 individuals aged 18-45. In the studied adult population, women are more prone to depression than men, and those who exercised regularly had lower rates of depression than those who did not and physical exercise is significantly correlated with a lower frequency of anxiety and depression (Khanzada et al., 2015). In conclusion, we proposed a hypothesis that Physical exercise had a positively predictive effect on depression of college students (H1).

To better investigate the effect of physical exercise on the depression of college students, self-concept was introduced as a mediator based on a literature review. Self-concept refers to a person's view of himself or herself, which can also be defined as the sum of everyone's perception of himself or herself, and plays a guiding role in in everyone's mental function (Simons et al., 2012). Abundant research results have proven the positive correlation between participation in programs of physical activity and higher levels of physical self-esteem and global self-concept (Perikles, 2006). Several studies show that the influence of self-concept, physical exercise satisfaction, quality of life, and self-concept are related, that is, the university sports exercise satisfaction is higher, the higher the quality of life and self-concept. According to a cognitive theory of depression, if individuals have low self-concept clarity, their cognitive abilities are reduced and they develop or maintain depression (DING Qian, 2016). At the same time, empirical studies have shown that college students' self-concept has a negative predictive effect on depression (Fan and Yuan, 2001; Li, 2004). Thus, we hypothesized that self-concept is the mediating variable between physical exercise and depression in college students (H2).

How does college students' self-concept affect their depression? By reviewing relevant literature, this study intends to use social support as a mediator between depression and self-concept in college students. Social support refers to beneficial interpersonal relationships from friends, family members, other contacts, or with a larger society. Social support is more about generating a feeling that a person is cared for and loved, respected and valued, or is part of a network of mutual interpersonal commitment. Empirical studies show that individuals with low self-concept clarity are difficult to change peer support, and social support and self-concept are significantly positively correlated (Emery et al., 2015; Quinones and Kakabadse, 2015). According to the self-categorization theory of depression cognition, depression can only be affected when the corresponding depression cognition interacts with a specific interpersonal relationship, that is, the interaction between negative cognition and negative interpersonal relationship can predict depression (Abramson, 1989). A good amount of empirical studies have also found that self-concept not only has a direct effect on depression (Wang and Qian, 2010), but also indicates the existence of other mediating variables, among which social support is one of the influencing factors. Therefore, we hypothesized social support might mediate the relationship between self-concept and depression (H3).

According to the interaction theory of social symbols, society is composed of interacting individuals who construct corresponding social roles, relationships, and standards. Physical exercise is a social interaction involving not only individuals but also groups (Mead, 1934). Some studies have pointed out that the interaction between individuals, family, and friends can promote the acquisition of the instrumental support, help maintain the level of physical exercise and active participation, so as to change the lifestyle and improve the level of mental health (Semmer et al., 2008). A large number of studies have shown that exercise can relieve depression and improve adverse health effects; at the same time, the study found that in the process of physical exercise, combined with their own interest and hobby to interact with different people, virtually promotes its own social ability, friendship, support, understanding social support, so as to enhance the self-worth and self-esteem, which reduces negative emotions, such as depression (Tang et al., 2007; Xu, 2011). Furthermore, Maslow's hierarchy of needs theory believes that friends, family, and other communication are the basic needs of belonging and communication. When an individual's needs are not being met, he or she will have negative emotions, such as anxiety and depression. Therefore, when the needs for interpersonal communication are not met, individuals will get less social support, which will easily lead to depression (Maslow, 1964). Therefore, this study hypothesized that social support is a mediating variable between physical exercise and depression in college students (H4).

In conclusion, this study intends to explore the influence mechanism of physical exercise on college students' depression, and construct a chain mediation model of physical exercise on college students' depression with social support and self-concept as mediating variables. The hypothesis model is shown in **Figure 1**.

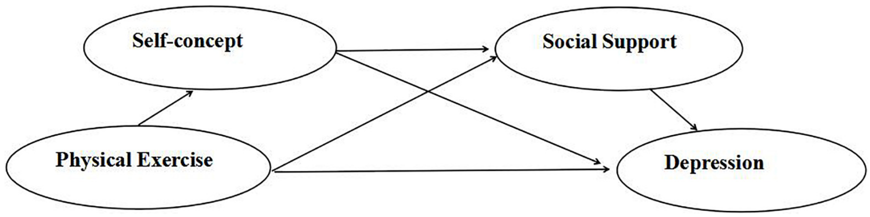


FIGURE 1 | The chain-mediated modle of self-cocept and social support.

MATERIALS AND METHODS

Study Design and Participants

Participants were recruited from undergraduates at some universities located in Jiangxi province, China. Due to the impact of COVID-19, the researchers asked the participants to answer an online questionnaire. Among the collected questionnaires, 1097 of which were valid, with an effective rate of 90%, with 440 male students accounting for 41 percent and 639 female students accounting for 59 percent. The subjects ranged in age from 17 to 24 (21.53 ± 0.84).

Measurements

Depression Scale

This study adopted the Self-Rating Depression Scale (the version by Zhang (1997)). This consists of 20 items using the 4-point Likert scale ranging from "no or little time", to "most or all time". The lowest score is 0, and the highest score is 60. The higher the score is, the more serious the degree of depression. In the actual measurement, the internal consistency coefficient of the questionnaire is 0.83.

Physical Activity Scale

The Physical Activity Scale was used to measure the level of the participants' physical activity. It was based on the Physical Exercise Intensity Scale for College Students (Liang, 1994). The scale was divided into three dimensions with 14 items: Intensity, Duration, and Frequency of Exercise. Higher scores indicate more intense exercise. In the current study, the scale's Cronbach's alpha was 0.86.

Social Support Scale

The Social Support Comprehension Scale adapted by Jiang Ganjin was adopted in this study,which consists of 12 items. The scale contains three dimensions:Family Support, Friend Support and Others. Respondents agreed to use a seven-point Likert scale that ranged from "very dissatisfied" to "dissatisfied." In the study, the final scale's Cronbach's alpha was 0.87.

Self-Concept Scale

Self-concept was evaluated by the self-concept scale, which was developed by Zhou based on the Self-concept clarity scale (Niu et al., 2016). The questionnaire was scored on a 7-point Likert scale ranging from 1," very dissatisfied "to 7," dissatisfied". Cronbach's coefficient of this measure in this study was 0.90.

Data Analysis

SPSS24.0 and macro program PROCESS were used to sort out and analyze the data for depression, physical activity, social support and self-concept. SPSS was first used for correlation analysis, and then PROCES model 6 was used to analyze the mediation effect.

RESULTS

Common Method Biases

Since the data in this study were filled in by the anonymous self-report method of the subjects, there may be a common method bias, so the Harman single-factor test method was used to test the data for bias. The factors with eigenvalues greater than 1 were extracted by testing, and it was found that there were 11 factors greater than 1, and the variation explained by the first factor was 22.88%, which was less than 40% of the critical value standard, indicating that there was no serious common method deviation in the data in this study.

Descriptive Statistics and Correlation Analyses

Table 1 reports descriptive statistics and correlative statistics among variables. Physical exercise was negatively correlated with depression ($r = -0.221$, $p < 0.01$), and was positively related to self-concept ($r = 0.291$, $p < 0.05$), social support ($r = 0.1611$, $p < 0.01$).

TABLE 1 | Pearson correlation analysis among different variables.

	M ± SD	1	2	3	4
physical exercise	8.04 ± 2.47	1			
depression	37.14 ± 7.77	−0.221**	1		
social support	62.37 ± 13.01	0.161**	−0.529**	1	
self-concept	48.92 ± 9.56	0.291**	−0.326**	0.143**	1

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

TABLE 2 | Bootstrap mediating effects of social support and self-concept.

Paths	Boot SE	Boot LLCI	Boot ULCI	Effect of SUM
PE - > DS	−0.08	−0.13	−0.02	
PE - > SC- > DS	−0.103	−0.18	−0.05	0.37
PE - > SS- > DS	−0.070	−0.10	−0.03	0.25
PE - > SC- > SS- > DS	−0.023	−0.03	−0.003	0.08
Total Effect	−0.196	−0.20	−0.11	

PE, physical exercise; DS, depression; SC, self-concept; SS, social support.
P < 0.05,P < 0.01 and ***P < 0.001.*

Mediation Analyses

In order to better explore the mechanism of physical exercise on depression of college students, self-concept and social support were considered to be the mediators.We use Model 6 of PROCESS to test the mediating effect.

Based on the structural equation, the chain mediating effect of self-concept and social support was further tested by Bootstrap (repeated sampling 5,000 times) at the deviation corrected percentile. As presented in **Figure 1** and **Table 2**, physical exercise was a direct positive predictor of self-concept ($\beta = -0.38$, $P < 0.001$), self-concept negatively predicted depression ($\beta = -0.27$, $p < 0.001$), and the mediating effect of self-concept on physical exercise on depression was 0.103, so H1 was supported. Physical exercise significantly positively predicted social support ($\beta = -0.13$, $P < 0.01$), social support negatively predicted depression ($\beta = -0.54$, $P < 0.001$), and the mediating effect of social support on the effect of physical exercise on depression was 0.07, thus H2 was established. At the same time, self-concept positively predicted social support ($\beta = -0.11$, $P < 0.01$), and the mediating effect of self-concept and social support on the effect of physical exercise on depression was 0.023, and the results also showed that physical exercise had a negative predictive effect on the depression of college students ($\beta = -0.08$, $P < 0.05$), the direct effect was 0.08, so H3 was valid, which indicates the mediating role of self-concept and social support. It can be seen that self-concept and social support play a part of mediating chain role in the effect of physical exercise on depression of college students.

DISCUSSION

A cross-sectional study was conducted to understand the relationship between physical exercise, depression, self-concept, and social support, and analyze the chain mediating role of self-concept and social support.

The Influence of Physical Exercise on Depression of College Students

This study shows that physical exercise can negatively predict the depressed mood of college students, which is obviously consistent with the previous result that physical exercise is negatively correlated with the risk of depression (Edman et al., 2014). Therefore, we should improve the physical exercise level of college students to reduce depression and other negative emotions of college students (Ruby et al., 2011). Empirical studies have shown that by testing the time, intensity, and frequency of college students' physical exercise, moderate to large amounts of physical exercise have a great effect on alleviating negative emotions, that is, college students with moderate to large amounts of physical exercise have a lower level of anxiety and depression [36]. Studies have also shown that both aerobic exercise and traditional physical activities have positive effects on the improvement of individual depression (Rimer et al., 2012). Therefore, we should attach great importance to the impact of physical exercise on college students' depression.

Mediating Effects of Self-Concept and Social Support

This study shows that self-concept and social support play a chain intermediary role in the relationship between physical exercise and depression, indicating that the mediating effect of self-concept and social support was established. In the model of the study, physical exercise can not only directly affect the college students' levels of depression, but also indirectly, by self-concept and social support, affect college students' depression.

It can be concluded that physical exercise could alleviate college students' depression mood and their self-concept cognition by means of physical exercise, so as to obtain greater social support and reduce the level of depression. Martinez pointed out in his study that the level of individual self-concept can indeed improve social perception through physical exercise, to better integrate into a peer group, and thus obtain more social support and alleviate depression (Yong and Tian, 2019). Many previous studies have also found that through physical exercise one can obtain more social support, relieve stress levels, and reduce adverse emotions, such as anxiety and depression (Salmon, 2001). Hence, it is necessary to guide our students to engage in more physical exercise, and to actively improve their self-concept awareness and social support level, so as to reduce their depression.

Implications

Physical exercise can positively predict college students' depression and improve their positive emotions through appropriate physical activities. The mediating effect of self-concept and social support was also found in our study. Therefore, in addition to actively improving the level of self-cognition of college students, it is necessary to improve their interpersonal communication skills, so as to reduce the level of depression.

Limitations and Future Studies

We can still find deficiencies in the study, which need to be improved in the future research. This study only discusses the influence of physical exercise on depression from the perspective of psychology, which can be discussed with the method of neuroscience in future research.

CONCLUSION

This study shows that physical exercise has a negative effect on depression in college students' depression,and social support and self-concept played a chain mediating role between physical exercise and depression. The mediating effect of self-concept and social support on depression indicates that the two variables have important practical significance in improving college students' mental health. The study will provide a reference for educators to improve the healthy physique and psychological quality of college students in the future.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JZ and SZ conceived and designed the experiments and wrote the manuscript. JZ, SZ, and ZH carried out the protocol and the questionnaire survey and revised the manuscript. JZ analyzed the data. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the Jiangxi University Humanities and Social Science Research Project (XL1418).

REFERENCES

Abramson, L. Y. (1989). Hopelessness depression: a theory- based subtype of depression. *Psychol. Rev.* 90, 358–372. doi: 10.1037/0033-295x.96.2.358

Crone, D. (2003). Physical activity and mental health. *Life Day* 7, 23–28. doi: 10.1108/13666282200300017

Edman, J. L., Lynch, W. C., and Yates, A. (2014). The impact of exercise performance dissatisfaction and physical exercise on symptoms of depression among college students: a gender comparison. *J. Psychol.* 148, 23–35. doi: 10.1080/00223980.2012.737871

Emery, L. F., Walsh, C., and Slotter, E. B. (2015). Knowing who you are and adding to it: reduced self-concept clarity predicts reduced self-expansion. *Soc. Psychol. Pers. Sci.* 6, 259–266. doi: 10.1177/1948550614555029

Esmaeilzadeh, S. (2015). The association between depressive symptoms and physical status including physical activity, aerobic and muscular fitness tests in children. *Environ. Health Prev. Med.* 20, 434–440. doi: 10.1007/s12199-015-0484-0

Fan, F., and Yuan, J. (2001). The relationship between self-concept and mental health of college Students. *Chin. J. Ment. Health* 2, 76–77. doi: 10.3321/j.issn:1000-6729.2001.02.002

Ibrahim, A. K., Kelly, S. J., Adams, C. E., and Glazebrook, C. (2013). A systematic review of studies of depression prevalence in university students. *J. Psychiatr. Res.* 47, 391–400. doi: 10.1016/j.jpsychires.2012.11.015

Khanzada, F. J., Soomro, N., and Khan, S. Z. (2015). Association of physical exercise on anxiety and depression amongst adults. *J. Coll. Physicians Surg. Pak.* 25, 546–548.

Li, J. (2004). The relationship between self-concept and depression in college students. *J. Jining Med. Coll.* 27:2. doi: 10.3969/j.issn.1000-9760.2004.04.019

Liang, D. (1994). Stress level of college students and its relationship with physical exercise. *Chin. J. Ment. Health* 8:2.

Maslow, A. H. (1964). Toward a psychology of being. *Philos. Phenomenol. Res.* 25:288. doi: 10.1037/10793-003

Mead, H. G. (1934). *Mind, Self, and Society*. Chicago Press, IL: University of Chicago Press.

Morgan, W. P., and Ellickson, K. A. (1989). "Health, anxiety, and physical exercise," in *Anxiety in Sports: An International Perspective*, eds D. Hackfort and C. D. Spielberger (New York: Taylor & Francis), 165–182. doi: 10.4324/9781315781594-14

Niu, G., Sun, X., Zhou, Z., Tian, Y., Liu, Q., and Lian, S. (2016). The effect of social networking site use on self-concept clarity among adolescents: the mediating role of social comparison. *Psychol. Sci.* 39, 6. doi: 10.16719/j.cnki.1671-6981.20160115

Perikles, K. (2006). The role of sport and physical education in self-concept development of children and adolescents. *Inq. Sport Phys. Educ.* 18, 11–12.

Qian, D. (2016). Relationship between social network site use and depression in adolescents: multiple mediating effect of social comparison and self-concept clarity. *Chin. J. Clin. Psychol.* 24, 823–827. doi: 10.16128/j.cnki.1005-3611.2016.05.012

Quinones, C., and Kakabadse, N. K. (2015). Self-concept clarity, social support, and compulsive internet use: a study of the US and the UAE. *Comput. Hum. Behav.* 44, 347–356. doi: 10.1016/j.chb.2014.11.019

Rimer, J., Dwan, K., Lawlor, D. A., Greig, C. A., McMurdo, M., Morley, W., et al. (2012). Exercise for depression. *Cochrane Database Syst. Rev.* CD004366. doi: 10.1002/14651858.cd004366.pub5

Ruby, M. B., Dunn, E. W., Perrino, A., Gillis, R., and Viel, S. (2011). The invisible benefits of exercise. *Health Psychol.* 30, 67–74. doi: 10.1037/a0021859

Salmon, P. (2001). Effects of physical exercise on anxiety, depression, and sensitivity to stress: a unifying theory. *Clin. Psychol. Rev.* 21, 33–61. doi: 10.1016/S0272-7358(99)00032-X

Semmer, N. K., Elfering, A., Jacobshagen, N., Perrot, T., Beehr, T. A., and Boos, N. (2008). The emotional meaning of instrumental social support. *Int. J. Stress Manage.* 15, 235–251. doi: 10.1037/1072-5245.15.3.235

Simons, J., Capio, C. M., Adriaenssens, P., Delbroek, H., and Vandenbussche, I. (2012). Self-concept and physical self-concept in psychiatric children and adolescents. *Res. Dev. Disabil.* 33, 874–881. doi: 10.1016/j.ridd.2011.12.012

Slykerman, R. F., Thompson, J. M. D., Coomarasamy, C., Wall, C. R., and Mitchell, E. A. (2019). Early adolescent physical activity, sleep and symptoms of depression at 16 years of age. *Acta Paediatr.* 109, 1394–1399. doi: 10.1111/apa.15140

Sun, Y. L., Wang, J., Yao, J. X., Ji, C. S., and Jin, Y. H. (2014). [Physical exercise and mental health: cognition, anxiety, depression and self-concept]. *Prog. Physiol.* 45, 337–342. doi: 10.1057/9780230375918

Tang, W., Wang, X., and Zheng, G. (2007). The effect of sports on depression in college students. *J. Beijing Sport Univ.* 30:2. doi: 10.3969/j.issn.1007-3612.2007.06.031

Walker, E. R., Mcgee, R. E., and Druss, B. G. (2015). Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. *JAMA Psychiatry* 72, 334–341. doi: 10.1001/jamapsychiatry.2014.2502

Wang, Q., and Qian, M. Y. (2010). The influence of self-concept shame on college students' depression. *Chin. J. Sch. Health* 31:3.

Wu, X., Kirk, S. F., Ohinmaa, A., and Veugelers, P. (2016). Health behaviours, body weight and self-esteem among grade five students in Canada. *SpringerPlus* 5:1099. doi: 10.1186/s40064-016-2744-x

Xu, X. (2011). The influence of physical exercise on interpersonal communication problems of vocational secondary school students. *J. Jiangsu Ocean Univ.* 9:3. doi: 10.3969/j.issn.1008-3499.2011.11.046

Yong, W., and Tian, X. (2019). Research on the relationship between physical exercise and self-concept. *Sports* 121–122. doi: 10.3969/j.issn.1674-151x.2019.02.059

Zhang, M. Y. (1997). Positive and negative syndrome scale and its application. *J. Clin. Psychiatry* 7:3.

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

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

frontiers | Frontiers in Education

ORIGINAL RESEARCH
published: 15 June 2022
doi: 10.3389/feduc.2022.893237

Identification of Cognitive Activities That Underlie Variations in Lecture Note-Taking: An Exploration of Japanese and Chinese High School Students' Strategies in Mathematics Class

Mengsi Liu[1]* *and Yuri Uesaka*[2]

[1] Department of Educational Psychology, Graduate School of Education, The University of Tokyo, Bunkyo-ku, Japan,
[2] Center for Research and Development on Transition From Secondary to Higher Education/Graduate School of Education (Concurrent Staff), The University of Tokyo, Bunkyo-ku, Japan

OPEN ACCESS

Edited by:
María Soledad Segretin, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Aloysius H. Sequeira, National Institute of Technology, Karnataka, India
Joseph R. Boyle, Temple University, United States

***Correspondence:**
Mengsi Liu
mengsi@g.ecc.u-tokyo.ac.jp

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

Received: *10 March 2022*
Accepted: *26 May 2022*
Published: *15 June 2022*

Citation:
Liu M and Uesaka Y (2022) Identification of Cognitive Activities That Underlie Variations in Lecture Note-Taking: An Exploration of Japanese and Chinese High School Students' Strategies in Mathematics Class. Front. Educ. 7:893237. doi: 10.3389/feduc.2022.893237

Lecture note-taking has been proven beneficial for learning at different educational levels. Previous studies have largely focused on the relationship between the outcomes of note-taking on a blank paper (e.g., measurements of the quantity and/or quality of notes taken) and student learning performance. However, there is no consensus as to what makes good notes. It is difficult to judge whether lecture note-taking is effective based only on the measurements of the notes. Past explorations have not adequately considered the cognitive activities that accompany such a process. Thus, using the interview method, the present study aimed to identify how lecture note-taking is used as a cognitive activity, and what factors influence it. To increase the possible range of note-taking approaches that could be observed, data from different cultural environments in Japan and China were sampled. Semi-structured interviews were conducted with 20 high school students from both countries (10 in each) to explore the cognitive activities in which students engage when taking lecture notes in mathematics class. Based on learning strategy models and studies, as well as using a thematic analysis, a new hierarchical framework of lecture note-taking, comprising shallow and deep lecture note-taking, was proposed. Deep lecture note-taking uses cognitive, metacognitive, and resource management functions. Furthermore, a comparison of students from the two countries revealed that their beliefs and teachers' instructions were potential factors influencing their lecture note-taking. Utilizing interview as the research method allowed us to obtain new insights into the cognitive activities that accompany lecture note-taking, such as the metacognitive function, which has rarely been explored in previous research. Future work is expected to commence on new measures based on this theoretical framework that gauges the cognitive activity of lecture note-taking. This study also calls for the exploration of effective note-taking instruction that considers the cognitive activity of note-taking.

Keywords: lecture note-taking, cognitive activities, interview, deep note-taking, information processing

INTRODUCTION

Lecture note-taking is a ubiquitous student activity from the elementary to the university level. It is defined as an activity that includes writing on learning materials, underlining, highlighting, and marking important points in class. Lecture note-taking is an essential activity that promotes learning (Peper and Mayer, 1986; Christopoulos et al., 1987; Peverly et al., 2003). Theories on note-taking propose two reasons why note-taking is beneficial for learning (Di Vesta and Gray, 1972). On one hand, according to the encoding perspective, note-taking is the act of recording information from a lecture, which benefits students through their engagement at a deeper processing level. On the other hand, according to the external storage perspective, the process of note-taking allows students to review, process, and commit information to memory. In this study, we were interested in how students' note-taking affected their understanding of lectures while receiving knowledge in the classroom. Hence, this study focused on the encoding function of note-taking.

Studies have pointed out that the encoding function of note-taking is effective because it encourages the processing of information (e.g., Kiewra, 1985). To evaluate the effectiveness of this encoding function, studies have generally compared the test performance of a note-taking group that could not review their notes with a listening-only group that did not take notes. Kiewra (1985) qualitatively analyzed 56 studies on the efficacy of encoding, from which there were 33 studies that displayed effects in favor of note-taking. Moreover, a meta-analysis of the encoding effect by Kobayashi (2006) found that, compared to listening only, note-taking without review had a small effect on test performance ($d = 0.22$).

However, taking notes does not necessarily lead to understanding. For example, Mueller and Oppenheimer (2014) reported that students who simply transcribed the presented content verbatim had inhibited learning. In addition, compared with university students, secondary students may act in a less self-regulated manner, as they rely on the teacher's instructions and supplement their notes on their own only minimally (Yokoi, 1999). Moreover, in Shiba's (2018) tutoring, the student being instructed always focused on the formality of the notes; for instance, "taking neat notes is studying," rather than the content to be written down. Thus, although lecture note-taking is vital in the classroom, the effect on promoting learning may depend on how it is used. Furthermore, considering that the process of note-taking affects understanding of information, it would be appropriate to also consider that there are shallow and deep levels in lecture note-taking. For example, repeated transcribing might be a shallow level of note-taking as it does not contribute to understanding or classroom learning.

Lecture note-taking is also a critical learning strategy for developing self-regulated learning at the secondary level (Zimmerman et al., 1996). From the perspective of self-regulated learning, in addition to deep and shallow levels in lecture note-taking, metacognitive aspects such as monitoring might also be important. Self-regulated learning is considered a proactive process through which students acquire academic skills—setting goals, selecting and deploying strategies, and self-monitoring their effectiveness (Zimmerman, 2008)—which are regarded as valuable skills learned during school years and beyond (cf. lifelong learning; OECD iLibrary, 2015). The basic process of self-regulated learning is viewed as a three-stage cycle: forethought, performance, and self-reflection (Zimmerman and Schunk, 2011). To complete the cycle, students need to monitor their own cognitive processes and then adjust/control them. For example, while taking lecture notes, it might be vital to monitor their comprehension, mark what they do not understand during class, and use the notes taken by them to adjust learning activities after class. Notably, previous studies concerning effective note-taking have only focused on what is written, that is, the results of note-taking. Thus, the cognitive and metacognitive aspects of lecture note-taking have not been examined. The present study focuses on the encoding function of note-taking, which is regarded as promoting students' understanding. In addition, we examined lecture note-taking on the axes of cognitive aspects, which consist of deep and shallow levels, and metacognitive aspects to grasp the whole figure of lecture note-taking from the viewpoint of cognition and metacognition.

Previous studies have measured the effects of lecture note-taking on learning by simply measuring the note quantity. Many studies have shown that the more notes students write in class, the better they perform on tests like fact recall and concept application tests (e.g., Kiewra and Benton, 1988; Peverly et al., 2007; Mueller and Oppenheimer, 2014). However, from this point of view, these studies could not explain why taking more notes verbatim had a negative relationship with the test performance of participants (Mueller and Oppenheimer, 2014).

Further, lecture note-taking is measured in terms of the quality of the notes; however, its criteria are varied. Numerous studies have examined the impact of note quality on memory of lecture content (Peverly et al., 2007, 2013; Peverly and Sumowski, 2012; Nakayama et al., 2017). Previous studies have shown that high-quality note-taking positively affects memory tests; however, the criteria for note quality vary across studies (e.g., Peverly and Sumowski, 2012).

Nevertheless, the critical issue is that cognitive processes of lecture note-taking are not considered. Both quantity and quality of notes are examined only in terms of the results of lecture note-taking; however, what the learner was thinking at the time is ignored, even though it is essential. For instance, even if the students write the same note, it is difficult to determine whether they just copied it verbatim or took it because they thought it was important. That is, there is a possibility that even if the same note contents are written, the learning effect would be different because of the different cognitive processes involved. However, without considering this, the true quality of lecture note-taking cannot be examined.

For example, in the study of learning strategies, the axes of shallow and deep strategies have been proposed and discussed (e.g., Marton and Säaljö, 1976; Murayama, 2007). We consider that lecture note-taking should exist these deep and shallow level use axes. Previous studies have assessed the quality of cognitive activity along these axes. For instance, Marton and Säaljö (1976) interviewed university students to understand their learning processes and discovered two learning strategies: surface-level and deep-level processing. In addition, learners who use deep-level processing strategies performed better than those who use surface-level processing strategies (e.g., Ramsden, 1988). However, these aspects of lecture note-taking have not been examined thus far. For example, copying down might be shallow processing of note-taking, while understanding the meaning or clarifying what is not understood might be deep processing. Nevertheless, it is unknown what the specifics are.

This study aimed to investigate cognitive activities that underlie variations in lecture note-taking. This study sampled data from different cultural environments in Japan and China to increase the possible range of note-taking approaches that could be observed. For example, on the one hand, in Japan, there is a strong cultural emphasis on the teachers' practices regarding writing on the blackboard and note-taking instructions (e.g., Ohtsubo and Higashibata, 2012). In addition, the Japanese national curriculum refers to using lecture notes as an assessment of "independent attitude toward learning (Ministry of Education, Culture Sports Science and Technology, 2020)." On the other hand, note-taking tends not to be emphasized in China (Wang, 2014). Furthermore, the Chinese national curriculum makes no specific mention of lecture note-taking. Thus, in these different contexts regarding note-taking, we considered that Japanese and Chinese students might engage in different note-taking activities.

To investigate cognitive activities that underlie variations in lecture note-taking, this study focuses on lecture note-taking during mathematics classes. First, mathematics has been considered a critical area for self-regulation because many students face difficulty with it, and effective use of learning strategies can enhance their learning and achievement (Patricia et al., 2017). However, even university students might not be able to take notes efficiently in their mathematics class (Fukawa-Connelly et al., 2017). We consider that identifying how students take notes in math class is vital for subsequent educational practice. Further, in this study, we will also examine the factors influencing students' lecture note-taking use, such as deep-level note-taking. Our results could lead to fostering students' use of deep note-taking.

Students' behavior can be greatly influenced by their beliefs (Dunlosky et al., 2015). It was noted that their note-taking beliefs affected the actual use of the notes (e.g., Bonner and Holliday, 2006; Witherby and Tauber, 2019). For example, previous studies have reported that students take notes as they believe that it helps them understand the lesson (Van Meter et al., 1994). This is consistent with the perspective of the encoding function of note-taking. However, beliefs about the other storage functions—such as taking notes for later reviews—have not been addressed. Based on previous findings, we can predict that the beliefs students hold will influence whether they take notes, while its explicit effects remain unknown.

Furthermore, students' use of note-taking also appears to be influenced by external factors, such as teachers' instruction. For instance, Titsworth (2001, 2004) and Titsworth and Kiewra (2004) found that inserting organizational cues and providing explicit statements that a theory or sub-theme of the theory was about to be discussed in a lecture can lead to an increase in the quality and quantity of information in notes as well as better test performance. Moreover, it was reported that inserting pauses in a lecture allows students time to consult with their peers or instructors (Ruhl, 1996), and preparing copies of PowerPoint slides for students (e.g., Marsh and Sink, 2010; Williams et al., 2012) can enhance their learning. From these studies, we can infer that the instructions provided by teachers affect students' note-taking or directly affect their learning. However, the specifics are unknown; few studies have investigated the effects of the different types of note-taking instruction and their effect on students' actual use of note-taking.

To better explore how different types of note-taking instruction influence note-taking, we believed it would be informative to collect data from both Japan and China. Although both Japan and China are considered to have high academic achievement among students in East Asia, the instruction style for note-taking may differ between them. For example, most teachers in China did not emphasize how to take notes in class (e.g., Chen, 2000; Wang, 2014). However, teachers' blackboard instructions and note-taking are stressed in Japan (Ohtsubo and Higashibata, 2012). For example, Uosaki (2017) conducted a questionnaire survey of university students and reported that students had received teachers' instruction on items such as "writing in a way that allows for easy note review," "writing important words," and "writing a large amount of text" before. From these studies, we assumed that collecting data from both Japan and China would allow for a better exploration of the effects of different note-taking instructions.

The primary purpose of this study was to explore how lecture note-taking is used as a cognitive activity, as well as which attributes of students and what instructions from teachers would affect the use of note-taking. We conducted interviews with Japanese and Chinese high school students about their lecture note-taking uses in class. Thus, sampling data from different cultural environments was carried out to increase the possible range of note-taking approaches that could be observed. Furthermore, it allows for a better exploration of the underlying factors of students' note-taking. An investigation of these issues can give us valuable insights into why learners use note-taking less effectively and the crucial facet that lecture note-taking functions. In addition, strategies for note-taking instructions can be suggested to develop self-regulated learners in educational practice.

MATERIALS AND METHODS

A qualitative research design was used to explore the cognitive activities underlying students' various lecture note-taking, and the background factors that might influence it. We used semi-structured qualitative interviews and then applied thematic analysis, as proposed by Boyatzis (1998).

Ethics Approval

The study was approved by the Research Ethics Committee of the university to which the authors are affiliated. Its purpose, nature, and assurances on the confidentiality and anonymity of published data have been clearly stated. All participants were verbally briefed on the interview day about the kinds of questions they would be asked, and they signed appropriate consent forms.

Participants

Twenty high school students from five high schools in Japan and seven high schools in China participated in this study. The participants were from 10th to 12th grade (approximately 15–18 years old), who were selected using snowball sampling. **Table 1** provides an overview of the gender, grade, and school status of each of the 20 participants.

Procedure: Semi-Structured Interviews

In autumn 2019, semi-structured qualitative interviews, intended to investigate students' actual lecture note-taking use and their perceptions of it, were conducted with 10 participants from each country. The interviewees were asked to bring the notebooks or other materials that they used during their actual math class. A semi-structured interview guideline and a series of questions in a conversational style was designed and used; for example, "Could you tell me how you take notes in your regular math class?" and "Why did you take notes in that way?" All sessions were audiotaped and lasted for approximately 30 min.

Data Analysis

The first author immediately transcribed verbatim all the audiotapes after each session. An inductive thematic analysis was applied to the primary transcripts using the coding method proposed by Boyatzis (1998). All labels, categories and their definitions were recorded in Excel as a codebook. The interviews were analyzed in Japanese, and the excerpts that were used to illustrate our results in the current paper were translated into English.

First, the transcripts of each group—Japanese and Chinese—were coded. The transcript of Student 1 in the Japanese group was coded, and the derived codes were given labels. Next, Student 2's transcript was coded using these labels, and new labels were derived and defined. Then, the remaining transcripts were coded based on the accumulated labels. Similarities among the labels were sought. The Chinese group underwent the same procedures.

Next, the two groups were compared and contrasted. Within the subsample, we developed categories and subcategories, based on which the subgroups were compared. We also sequenced the categories and detailed the activities used by the participants during their lectures.

TABLE 2 | Primary materials for note-taking in the Japanese and Chinese groups.

	Notebook	Textbook	Total
Japanese students	9	1	10
Chinese students	4	6	10

Subsequently, after the initial coding, an external researcher—a native Chinese graduate student fluent in Japanese—who was unfamiliar with the study's aims independently analyzed 20% of the data. Her coding was compared with the authors' coding; this showed an inter-rater agreement of 79% (Gwet's AC1 was used). The disagreements were discussed, and codes were revised for inclusion in the analysis.

RESULTS

Typical Lecture Note-Taking Style of Japanese and Chinese High School Students

First, we confirmed the overall differences between the two groups' typical note-taking styles before conducting a more specific analysis. The lecture notes of nine out of the ten Japanese students were observed to be consistently taken in their notebooks (see **Table 2**). They seem to have been noted down exhaustively (see **Figure 1**). By contrast, more than half of the Chinese students' lecture notes were seen to be taken directly on their textbooks (see **Table 2**) rather than their notebooks, and the contents of the notes were inferred to have been jotted down selectively (see **Figure 2**). Next, a thematic analysis was adopted for coding and comparison to discuss the differences more specifically.

Categories Identified in the Current Study

As mentioned above, based on studies of learning strategies (e.g., Murayama, 2007) and the thematic analysis, students

TABLE 1 | Participants' gender, grade, and school characteristics.

Japanese group					Chinese group				
ID	Gender	Grade	Area	School status	ID	Gender	Grade	Area	School status
J1	Female	11th	Tokyo	Private	C1	Female	10th	Hunan	Private
J2	Female	11th	Tokyo	Private	C2	Male	10th	Jiangsu	Public
J3	Female	11th	Kanagawa	Public	C3	Female	10th	Hunan	Private
J4	Male	11th	Kanagawa	Public	C4	Male	11th	Hunan	Private
J5	Female	11th	Kanagawa	Public	C5	Male	11th	Hunan	Private
J6	Female	10th	Tokyo	Public	C6	Male	11th	Tianjin	Public
J7	Female	10th	Tokyo	Private	C7	Male	11th	Hunan	Public
J8	Female	10th	Tokyo	Private	C8	Male	12th	Hunan	Public
J9	Female	10th	Tokyo	Private	C9	Female	11th	Hunan	Public
J10	Female	10th	Tokyo	Private	C10	Female	11th	Hunan	Public

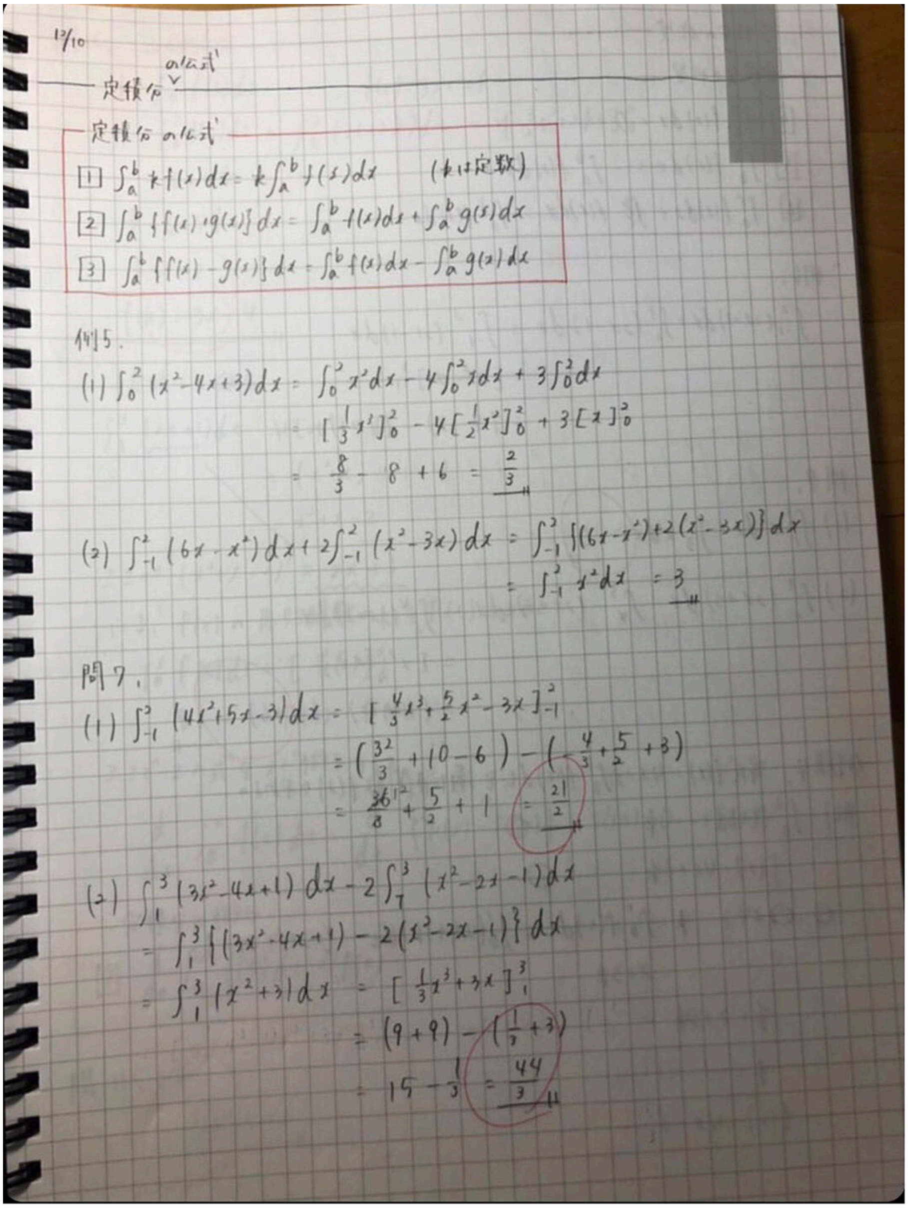
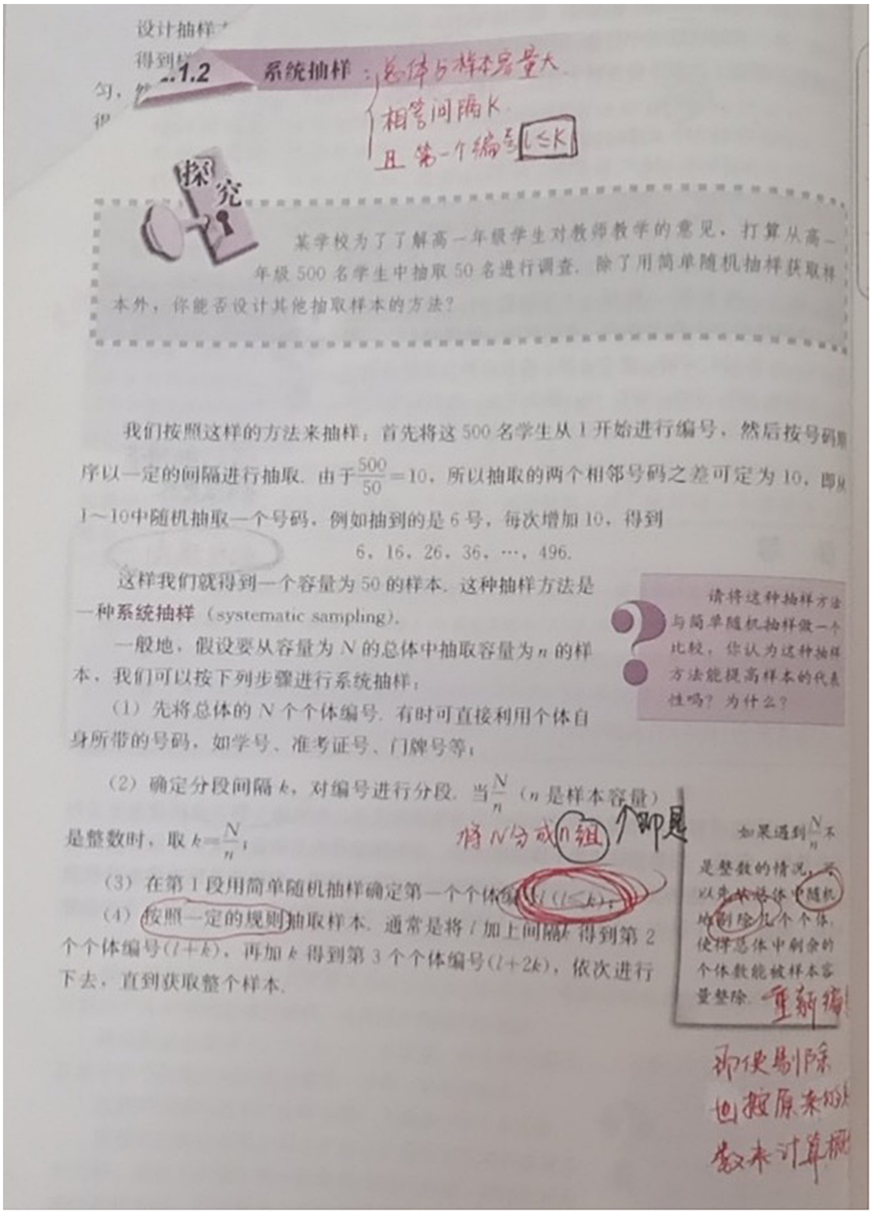


FIGURE 1 | Lecture notes in the Japanese group.

from both Japan and China identified two categories of lecture note-taking—shallow lecture note-taking and deep lecture note-taking. Deep lecture note-taking was further divided into three subcategories: cognitive, metacognitive, and resource management functions (see **Figure 3**).

Shallow Lecture Note-Taking

Based on previous studies (e.g., Murayama, 2007), lecture note-taking that relied on passive activities to help in the recall of information, as reported by the students, was categorized as shallow lecture note-taking: (a) copying the board, (b) acting as per instructions, (c) unconscious use of color, and (d) copying the textbook. Regarding copying from the board, students reported that they copied what the teacher wrote on the board during math class. For example, a student said, "The board: first, I definitely would copy the teacher's notes on the board. In the case of a formula, it usually has a detailed description of the calculation process and so on; I would copy all of them down." Acting as per instructions was coded by students, mentioning that they wrote down as the teachers told them to. About the unconscious use of color, students explained that they take notes as colorfully as they like, regardless of the content. One student said, "Maybe the color changes depending on how I am feeling that day." Finally, copying the textbook was coded by students who reported that they would copy diagrams and other information from the



2.1.2 系统抽样

探究

某学校为了了解高一年级学生对教师教学的意见，打算从高一年级500名学生中抽取50名进行调查．除了用简单随机抽样获取样本外，你能否设计其他抽取样本的方法？

我们按照这样的方法来抽样：首先将这500名学生从1开始进行编号，然后按号码顺序以一定的间隔进行抽取．由于$\frac{500}{50}=10$，所以抽取的两个相邻号码之差可定为10，即从1～10中随机抽取一个号码，例如抽到的是6号，每次增加10，得到

6，16，26，36，…，496．

这样我们就得到一个容量为50的样本．这种抽样方法是一种**系统抽样**（systematic sampling）．

一般地，假设要从容量为N的总体中抽取容量为n的样本，我们可以按下列步骤进行系统抽样：

（1）先将总体的N个个体编号．有时可直接利用个体自身所带的号码，如学号、准考证号、门牌号等；

（2）确定分段间隔k，对编号进行分段．当$\frac{N}{n}$（n是样本容量）是整数时，取$k=\frac{N}{n}$；

（3）在第1段用简单随机抽样确定第一个个体编号l（$l\le k$）；

（4）按照一定的规则抽取样本．通常是将l加上间隔k得到第2个个体编号（$l+k$），再加k得到第3个个体编号（$l+2k$），依次进行下去，直到获取整个样本．

请将这种抽样方法与简单随机抽样做一个比较，你认为这种抽样方法能提高样本的代表性吗？为什么？

如果遇到$\frac{N}{n}$不是整数的情况，可以先从总体中随机地剔除几个个体，使得总体中剩余的个体数能被样本容量整除．

FIGURE 2 | Lecture notes in the Chinese group. The mathematics textbook used was published by the People's Education Press.

textbook directly into their notes during class. For example, another student said, "While looking at the textbook, my teacher would say, 'Write here,' and I would copy those—some summarized detailed sentences in the textbook—into my notebook."

Deep Lecture Note-Taking

Contrastingly, lecture note-taking that required extra activities to improve understanding of the meaning (such as elaboration of information and evaluating information) was categorized as deep lecture note-taking, which was further divided into cognitive, metacognitive, and resource management functions of lecture note-taking.

Cognitive Function

The students reported several cognitive strategies in lecture note-taking for processing information from lectures; these were categorized as the cognitive function of lecture note-taking. These were: (a) Notes from verbal explanations, (b) paraphrasing difficult words, (c) teacher's supplement, (d) notes on what were considered important, (e) notes on the emphasis, (f) choosing what to write down, and (g) distinguishing content by color. Notes from verbal explanations refer to making supplemental notes regarding what their teacher is explaining verbally. Paraphrasing difficult words was explained as rephrasing mathematical terms and solutions into simple words to make them easier to understand. For example,

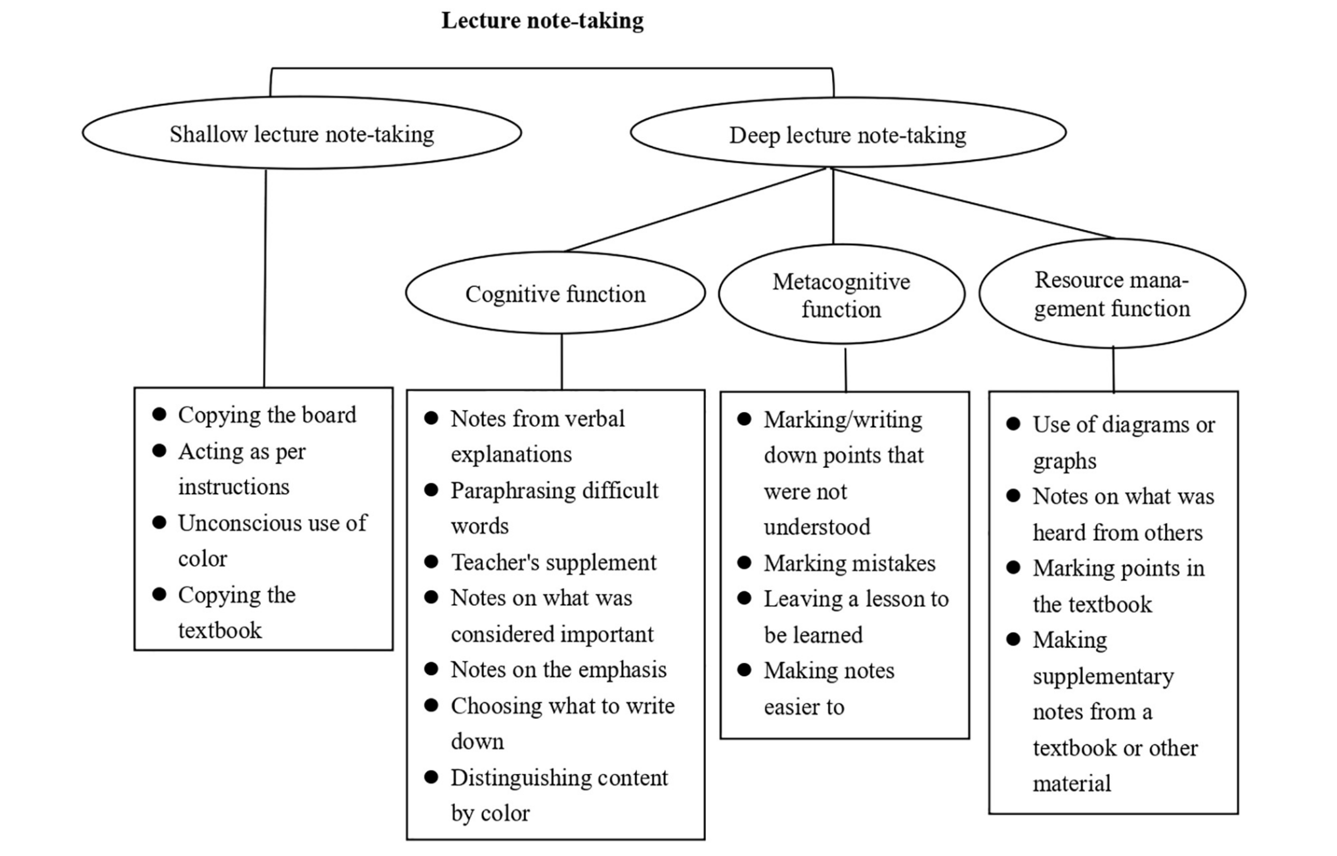


FIGURE 3 | A new framework for lecture note-taking.

"An easier way of reviewing is to break down the teacher's words into my own words so that it would not become too formal." The teacher's supplement shows that students take notes on things that are not in the textbook, such as alternate solutions to solve a problem provided by the teacher. For instance, one student stated, "The teacher gives us variants, and those are often used more. For example, like this one, which is not in the book at all, but they often show up on exam." Notes on what was important were coded by students stating that they mark what they think is essential. Notes on the emphasis convey that students usually take notes of what the teacher is emphasizing on. For example, "Also, these notes and stuff like that, where the teacher would draw a line that said it is important." The label of choosing what to write down signifies that they do not view all the information indiscriminately. Instead, they would selectively decide what to note. For example, one student said, "So, I only write the ones I do not understand, and the ones I do understand would be just written in the textbook." Finally, distinguishing content by color means that students use different color pens according to the content of their notes. For example, one student said, "The orange ones are teacher's notes, the ones written in pencil are my answers to the question, and the red ones are the correct answers to the questions."

Metacognitive Function

The second subcategory is metacognitive function of lecture note-taking that help students control and regulate their own cognition following four activities reported by the students: (a) marking/writing down points that were not understood, (b) marking mistakes, (c) leaving a lesson to be learned, and (d) making notes easier to understand.

First, marking/writing down points that were not understood was coded by the participating students as writing or marking down something they did not understand in the class. For example, "I did not understand this unit vector's meaning, so I added a question mark here." Marking mistakes means that they leave a reminder to themselves about the mistakes in the problems they solved. For example, a student said, "Umm. . .I would definitely mark the mistakes I made." Moreover, leaving a lesson to be learned conveys that students would leave a notice or lesson to prevent the same mistakes in the future. For example, one student stated, "I made a note of things like, "You need to be careful here." Making notes easier to understand signified the students' efforts to make their notes understandable from the perspective of others or their own perspective in the future. For example, "I am always conscious of making a notebook that could be readily understood by others as if I were showing it to someone else, and I have probably always been conscious of that for a long while." Further, one student stated that "If not (there are some markings in the notebook), I would feel confused about its extent. That is why I wrote it down."

Resource Management Function

The students also reported their regulatory strategies for controlling other resources besides their cognition; these were categorized as the resource management function of lecture note-taking. These were: (a) Notes on what was heard from others, (b) use of diagrams or graphs, (c) marking points in the textbook, and (d) making supplementary notes from a textbook or other material. Notes on what was heard from others refer to taking notes on the doubts they asked their teachers or friends to explain. For example, one student said, "If I did not understand what the teacher was saying, I asked my friend, who is a little bit smart, 'Why is this happening?,' and I would write down what he told me." The use of diagrams or graphs indicates that they are drawn to facilitate understanding of the concept or to think. For example, one student said, "It is easier to understand if I draw diagrams by myself, so I try to draw what I can in diagrams as much as possible in this way." Marking points in the textbook was coded by students stating that they underlined or marked essential concepts and points in the textbook. For example, one student said, "I would use a color pen, like a red pen, to circle it (considered important) with emphasis." Furthermore, the act of making supplementary notes from a textbook or other material means adding content from the textbook or other materials while taking notes in class.

Lecture Note-Taking Strategy Uses of Japanese and Chinese Students in the New Framework of Lecture Note-Taking

Overall, it was observed that both Chinese and Japanese students consistently exhibited two categories of lecture note-taking strategy, accompanied with cognitive activities: shallow lecture note-taking, and deep lecture note-taking consisting of cognitive, metacognitive, and resource management functions. We compared the categories and subcategories across the two groups (see **Table 3**) to explore their lecture note-taking precisely. In addition, **Figure 4** graphically presents the similarities and differences between the two groups.

In terms of shallow lecture note-taking, Japanese students reported more passive use of note-taking. Japanese students' lecture note-taking seemed to be more comprehensive, while that of Chinese students was more selective, which can be glimpsed from the materials they brought over to the interview. The total units of note-taking contents were 88 units for the Japanese students (M_j = 8.8) and 46 units for the Chinese students (M_c = 4.6). Thus, this implies that, in general, Japanese students took more detailed notes than their Chinese counterparts. Although several students in the two groups both reported that they simply copied what the teacher wrote on the board, only Japanese students reported their note-taking as both an "act as per instructions" and "copying the textbook." Two of the Japanese students mentioned that they wrote it down just as the teachers told them. For example, one student said, "The teacher said, "let us remember it and write it here!" so, I just wrote it like this." (J3)

TABLE 3 | Comparing the note-taking classifications between the two groups.

Categories	Japanese group	Chinese group
Subcategories	(Participant ID)	
Labels	*(J1-10)*	*(C1-10)*
Shallow lecture note-taking		
Copying the board	J1, J2, J3, J4, J5, J6, J7, J9	C3, C6, C9, C10
Acting as per instructions	J2, J3	
Unconscious use of color	J1, J3	C6
Copying the textbook	J1, J2, J7	
Deep lecture note-taking		
Cognitive function		
Notes from verbal explanations	J2, J3	C3, C4, C7, C10
Paraphrasing difficult words	J2, J3, J9	
Teacher's supplement	J3	C2, C3, C5, C7, C9
Notes on what was considered important	J1, J2, J3, J4, J7, J8, J9, J10	C1, C4, C7, C8, C9, C10
Notes on the emphasis	J1, J3, J5, J6, J9	
Choosing what to write down	J9	C3, C5, C6, C7, C8, C9, C10
Distinguishing content by color	J1, J2, J3, J9	
Metacognitive function		
Marking/writing down points that were not understood	J5	C2, C5, C9
Marking mistakes	J3, J4, J5, J6, J7	C1, C4, C10
Leaving a lesson to be learned	J2, J3, J5, J6	C8, C9
Making notes easier to understand	J1, J3, J6	
Resource management function		
Use of diagrams or graphs	J4, J6, J7, J10	C3, C5, C9
Notes on what was heard from others	J2, J4, J7	
Marking points in the textbook		C3, C4, C5, C10
Making supplementary notes from a textbook or other material	J5	

In addition, three of the ten Japanese students reported copying diagrams and other information from the textbook directly into their notes during class. For instance, "While looking at the textbook, my teacher would say, "Write here," and I would copy those—some summarized detailed sentences—in the textbook, into my notebook." (J2) Thus, as shown in **Figure 4**, although both Chinese and Japanese exhibited instances of passive note-taking, it might be more prevalent among Japanese students.

In terms of deep lecture note-taking, Chinese and Japanese students were relatively consistent in utilizing these cognitive, metacognitive, and resource management functions for their learning. Yet, specific differences were seen by comparing the two groups using thematic analysis. Concerning the cognitive function in lecture note-taking, only Japanese students reported their note-taking as "paraphrasing difficult words," "notes on the emphasis," and "distinguishing content by color." Meanwhile, we also found the metacognitive function in lecture note-taking in both groups; however, the activity "making notes easier to understand" was only reported by Japanese students. Furthermore, there were differences in the resource

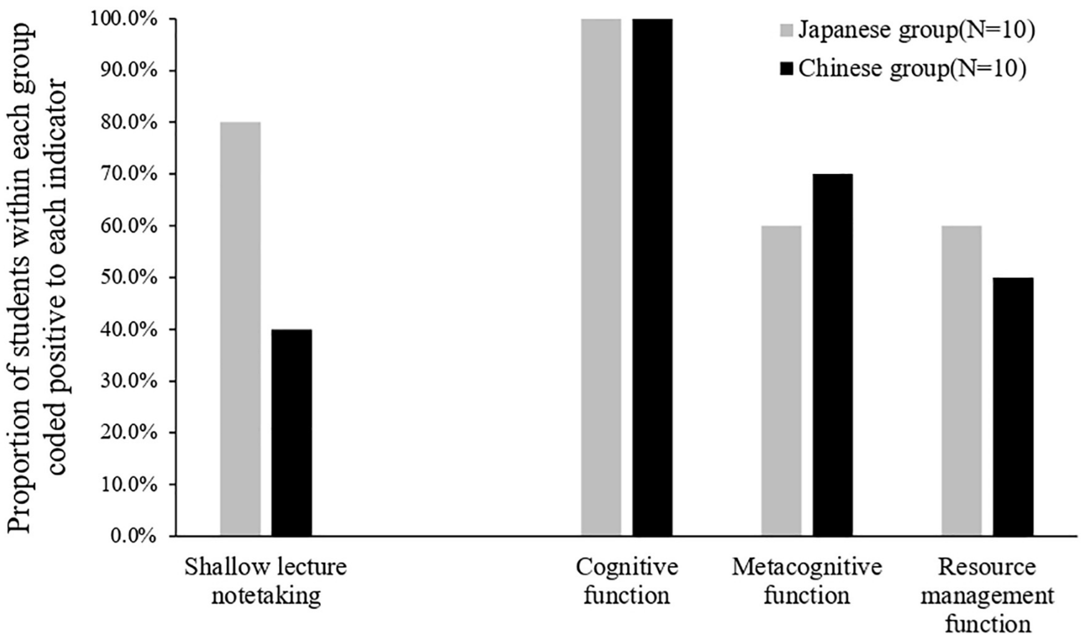


FIGURE 4 | Indicators of lecture note-taking activities.

management function. For instance, note-taking activities such as "notes on what was heard from others" and "making supplementary notes from a textbook or other material" were only found among Japanese students. In contrast, only Chinese students reported their note-taking as "marking points in the textbook."

Exploring the Background Factors Behind the Differences

The remaining thematic analysis data were used to explore the potential factors behind the differences in lecture note-taking between Japanese and Chinese high school students. The examination of each country's distribution across the beliefs and teachers' instruction of lecture note-taking involved in our research are shown in **Tables 4**, **5**.

Table 4 shows that, consistent with the encoding perspective of note-taking, 30% of students in both groups stated that they take notes because it helps them understand lecture content. For example, a student reported that "I also copy down the graph of a function, because it makes it easier for me to understand." (C3) Moreover, in line with the external storage perspective on note-taking, both Japanese and Chinese students mentioned that they take notes because doing so gives them material to review. However, the way in which notes are used for review differs. Approximately 90% of Japanese students reported taking lecture notes for reviewing later; for instance, "Because I would review it, I thought it would be better if it were easy for me to understand." (J10) While some students in the Chinese group took notes for simple reviewing too, a larger number of students mentioned that they were doing so for note-making after class. For example, "I wrote it down also for organizing my error book after class." (C9) In terms of the findings in **Table 4**, it can be inferred that these differences in the purpose of note review may cause the differences among the Japanese and Chinese high school students' lecture note-taking.

Meanwhile, **Table 5** indicates the effect of teachers' instruction on lecture note-taking as perceived by the two groups of students. First, only Japanese students stated that they took notes because the teachers gave explicit directions for writing. In addition, students of both groups reported that the teacher's cues on what is essential influence their lecture note-taking. Moreover, more students in Japan (vs. their Chinese counterparts) perceived the effect of the notes assessment (30%) and teachers' instruction (50%). In general, this implies that Japanese students are more aware of their teacher's instruction when taking lecture notes, which may be conducive to taking a few more notes.

DISCUSSION

This study applied the interview method to explore how lecture note-taking is used as a cognitive activity and the factors influencing the use of note-taking. Through thematic analysis, a new framework consisting of shallow and deep lecture note-taking was derived. Moreover, the interviews allowed us to clarify the specific characteristics of shallow and deep lecture note-taking. In addition, by comparing students from Japan and China, we inferred that the differences in lecture notes may result from the influence of the students' beliefs and teachers' instruction styles.

Regarding lecture note-taking as a cognitive activity, based on models of learning strategies studies and using thematic analysis,

TABLE 4 | Summary of the influence of beliefs on participants' note-taking.

	Lecture understanding		External storage			
			Simple review		Note-making after class	
	Number of students	Percentage	Number of students	Percentage	Number of students	Percentage
Japanese	3	30.0%	9	90.0%	9	10.0%
Chinese	3	30.0%	3	30.0%	4	40.0%

TABLE 5 | Summary of the influence of teachers' instruction on participants' note-taking.

	Teachers' instructions							
	Directions for what to write		Cues for what is important		Note assessment		Note instruction	
	Number of students	Percentage	Number of students	Percentage	Number of students	Percentage	Number of students	Percentage
Japanese	3	30.0%	5	50.0%	3	30.0%	5	50.0%
Chinese	0	0.0%	5	50.0%	1	10.0%	2	20.0%

a new framework of lecture note-taking comprising shallow lecture note-taking and deep lecture note-taking (cognitive function, metacognitive function, and resource management function) was proposed. In this new framework, some new aspects of lecture note-taking, such as the metacognitive function, were identified. This allowed us to go beyond and extend the previous discussion (e.g., Di Vesta and Gray, 1972; Kiewra and Benton, 1988; Nakayama et al., 2017) in note-taking.

Moreover, based on this new framework, similarities and differences in lecture note-taking between Japanese and Chinese students were also clarified. From a glimpse of the participants' actual notes, we found that the overall difference was that Japanese students take comprehensive notes in their notebooks while Chinese students take selective notes directly in their textbooks. Meanwhile, more specific characteristics between the two groups were found using the comparison method in the thematic analysis. Regarding shallow lecture note-taking, Japanese students reported more passive use of note-taking. On the other hand, although Chinese and Japanese students were relatively consistent in utilizing deep lecture note-taking for their learning, specific differences were seen in each function by comparing the two groups through the thematic analysis. For example, in terms of the resource management function, activities such as "notes on what was heard from others" were only reported by Japanese students. It can be inferred that there might be more interactive activities in the Japanese classroom, allowing students to ask others in class. Furthermore, only Chinese students addressed "marking points in the textbook." We could ascertain that textbooks may be used more often in a Chinese math class so that students take notes directly in the textbook.

Regarding which factor would affect the use of note-taking, we explored how students' perceptions of note-taking and beliefs about teachers' instruction might influence their note-taking. First, consistent with the encoding perspective of note-taking theory, both groups of students stated that they take notes because it helps them understand lecture contents (Van Meter et al., 1994; Bonner and Holliday, 2006). Nevertheless, in contrast to the previous study's finding that college students mentioned little about taking notes for review, high school students in our study explicitly talked about it, suggesting the need to examine note-taking at different stages. In addition, there were differences in the students' beliefs regarding lecture note-taking for review. Almost all Japanese students took notes for simple review, while some Chinese students took lecture notes to make their own notes after class, which may explain the abovementioned differences in their lecture note-taking.

However, in contrast to prior studies (e.g., Marsh and Sink, 2010), we found several aspects of instruction about note-taking—directions for what to write, cues for what is essential, note assessment, and note instructions—that influence lecture note-taking. Based on the differences between Japanese and Chinese students in lecture note-taking, it can be inferred that Japanese students are more aware of their teacher's explicit instructions on the notes' content and the assessment of their notes. This may lead to taking a few more notes, but not necessarily better ones. This study also explored different instructional approaches to note-taking, knowing that there are instructional approaches at the secondary level that are different from those at the university level; this gives us insights into the instructions that should be provided to improve student learning.

This study has several limitations and recommendations for follow-up research. Although the potential factors affecting note-taking have been discussed, there is no discussion on their specific effects. Second, the small sample size does not allow generalization to all secondary school students; thus, a larger sample of Japanese and Chinese students would allow for a more representative picture of secondary school students' lecture note-taking. Moreover, since this study explored note-taking activities in mathematics classes, whether the findings are transferable to other content area or subjects remains to be investigated. Furthermore, we also discussed the possible influences of teachers; however, the findings are derived from a student-based survey. Further evidence of this finding and new findings can be obtained by conducting teacher-based observation surveys. Last, in exploring the cognitive activities that underlie lecture note-taking, we did not assume the existence of gender differences and therefore did not fully balance the male and female participants. However, previous studies have illustrated gender differences in note-taking activities (e.g., Reddington et al., 2015). Therefore, caution may be needed in the interpretation of the framework proposed.

This study has several implications for educational practice. First, the present study pointed out the necessity to consider the cognitive activities involved in lecture note-taking. Future work is expected on new measures based on the proposed theoretical framework that gauge the cognitive activity of lecture note-taking. Meanwhile, whether the deep lecture note-taking effectively promotes learning requires empirical evidence. On the other hand, note-taking instruction may prompt students to take more notes; however, this does not necessarily have a positive effect on helping students understand the lecture. This study reveals that instructions promoting deep lecture note-taking should be emphasized. For example, the metacognitive function of lecture note-taking should be exploited more. In addition, encouraging students to utilize resources, such as textbooks for note-taking might be effective. These have important implications for reconsidering the need for note-taking instructions and the exploration of effective note-taking instruction that takes into account the cognitive activities of note-taking.

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 author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Tokyo/Office for Life Science Research Ethics and Safety. Written informed consent to participate in this study was provided by the participants.

AUTHOR CONTRIBUTIONS

ML conceived the present idea, carried out the implementation, performed data analysis, and took the lead in writing the manuscript. ML and YU contributed to the study design. Both authors provided critical feedback and helped shape the research, analysis and manuscript.

FUNDING

This research was supported by "17th Hakuhodo Research Grant for Child Education" from the Hakuhodo Foundation.

ACKNOWLEDGMENTS

We thank all participants in this study deeply. We also would like to express our gratitude to Affiliate Professor Shinichi Ichikawa for his guidance and helpful advice. Our thanks are also extended to Ms. Dongli Tan for her careful coding and suggestions.

REFERENCES

Bonner, J. M., and Holliday, W. G. (2006). How college science students engage in note-taking strategies. *J. Res. Sci. Teach.* 43, 786–818. doi: 10.1002/tea.20115

Boyatzis, R. E. (1998). *Transforming Qualitative Information: Thematic Analysis and Code Development*. London: Sage Publications.

Chen, J. (2000). Discussion on the guidance of students' lecture notes. *Shandong Educ. Res.* Z2:123. doi: 10.1152/advan.00126.2020

Christopoulos, J. P., Rohwer, W. D., and Thomas, J. W. (1987). Grade level differences in students' study activities as a function of course characteristics. *Contemp. Educ. Psychol.* 12, 303–323. doi: 10.1016/S0361-476X(87)80003-6

Di Vesta, F. J., and Gray, G. S. (1972). Listening And Note Taking. *J. Educ. Psychol.* 63, 8–14.

Dunlosky, J., Mueller, M. L., and Tauber, S. K. (2015). "The contribution of processing fluency (and beliefs) to people's judgments of learning," in *Remembering: Attributions, Processes, and Control in Human Memory: Essays in Honor of Larry Jacoby*, eds D. S. Lindsay, A. P. Yonelinas, and H. I. Roediger (New York, NY: Psychology Press), 46–64.

Fukawa-Connelly, T., Weber, K., and Mejía-Ramos, J. P. (2017). Informal Content and Student Note-Taking in Advanced Mathematics Classes. *J. Res. Math. Educ.* 48:567. doi: 10.5951/jresematheduc.48.5.0567

Kiewra, K. A. (1985). Providing the Instructor's Notes: an Effective Addition to Student Notetaking. *Educ. Psychol.* 20, 33–39. doi: 10.1207/s15326985ep2001_5

Kiewra, K. A., and Benton, S. L. (1988). The relationship between information-processing ability and notetaking. *Contemp. Educ. Psychol.* 13, 33–44. doi: 10.1016/0361-476X(88)90004-5

Kobayashi, K. (2006). Combined Effects of Note-Taking/-Reviewing on Learning and the Enhancement through Interventions: a meta-analytic review. *Educ. Psychol. Rev.* 26, 459–477. doi: 10.1080/01443410500342070

Marsh, E. J., and Sink, H. E. (2010). Access to handouts of presentation slides during lecture: consequences for learning. *Appl. Cogn. Psychol.* 24, 691–706. doi: 10.1002/acp.1579

Marton, F., and Säaljö, R. (1976). On qualitative differences in learning-II outcome as a function of the learner's conception of the task. *Br. J. Educ. Psychol.* 46, 115–127. doi: 10.1111/j.2044-8279.1976.tb02304.x

Ministry of Education, Culture Sports Science and Technology (2020). Shin Gakushu Shido Yoryo No Zemmen Jisshi to Gakushu Hyoka no Kaizen Ni tsu i te [Full Implementation of the New Courses of Study and Improvement of Learning Assessment]. Available online at: https://www.mext.go.jp/content/20201023_mxt_sigakugy_1420538_00002_004.pdf (accessed April 1, 2022).

Mueller, P. A., and Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: advantages of longhand over laptop note taking. *Psychol. Sci.* 25, 1159–1168. doi: 10.1177/0956797614524581

Murayama, K. (2007). "Gakushu horyaku: kodomo jishin no jiritsuteki na gakushu wo mezashi te [Learning Strategies: for children's self-regulated learning]," in *Zettai yakudatsu kyoiku shinri gaku*, ed. T. Fujita (Kyōto: Mineruvua Shoboō), 85–100.

Nakayama, M., Mutsuura, K., and Yamamoto, H. (2017). The possibility of predicting learning performance using features of note taking activities and instructions in a blended learning environment. *Int. J. Educ. Technol. High. Educ.* 14:6. doi: 10.1186/s41239-017-0048-z

OECD iLibrary (2015). *OECD Skills Outlook 2015: Youth, Skills and Employability*. Paris, Ile-de-France: OECD iLibrary.

Ohtsubo, and Higashibata. (2012). Writing on blackboard' by teachers and 'Note-taking' by students. Bulletin of the Faculty of Education, Kagoshima University. *Stud. Educ.* 63, 107–119.

Patricia, P. A., Dale, D. H., and Jeffrey, J. A. (2017). *Handbook of Self-Regulation of Learning and Performance.* New York, NY: Routledge.

Peper, R. J., and Mayer, R. E. (1986). Generative effects of note-taking during science lectures. *J. Educ. Psychol.* 78, 34–38. doi: 10.1111/j.1751-0813.2002.tb10969.x

Peverly, S. T., Brobst, K. E., Graham, M., and Shaw, R. (2003). College adults are not good at self-regulation: a study on the relationship of self-regulation, note taking, and test taking. *J. Educ. Psychol.* 95, 335–346. doi: 10.1037/0022-0663.95.2.335

Peverly, S. T., Ramaswamy, V., Brown, C., Sumowski, J., Alidoost, M., and Garner, J. (2007). What predicts skill in lecture note taking? *J. Educ. Psychol.* 99, 167–180. doi: 10.1037/0022-0663.99.1.167

Peverly, S. T., and Sumowski, J. F. (2012). What variables predict quality of text notes and are text notes related to performance on different types of tests? *Appl. Cogn. Psychol.* 26, 104–117. doi: 10.1002/acp.1802

Peverly, S. T., Vekaria, P. C., Reddington, L. A., Sumowski, J. F., Johnson, K. R., and Ramsay, C. M. (2013). The relationship of handwriting speed, working memory, language comprehension and outlines to lecture note-taking and test-taking among college students. *Appl. Cogn. Psychol.* 27, 115–126. doi: 10.1002/acp.2881

Ramsden, P. (1988). "Context and strategy: situational influences on learning," in *Learning Strategies and Learning Styles,* ed. R. R. Schmeck (New York, NY: Plenum Press), 159–184.

Reddington, L. A., Peverly, S. T., and Block, C. J. (2015). An examination of some of the cognitive and motivation variables related to gender differences in lecture note-taking. *Read. Writ. Interdiscip. J.* 28, 1155–1185.

Ruhl, K. L. (1996). Does nature of student activity during lecture pauses affect notes and immediate recall of college students with learning disabilities? *JPED* 12, 16–27.

Shiba, S. (2018). "). Improvement of learning ability by "Summary Notebook" to overcome learning difficulties," in *2018 Case Studies in Learning Support: Cognitive Counseling Practices for Developing Independent Learners (Excerpt)*, eds Y. Uesaka, S. Shiba, and S. Ichikawa (Japan: University of Tokyo), 36–52.

Titsworth, B. S. (2001). The effects of teacher immediacy, use of organizational lecture cues, and students' notetaking on cognitive learning. *Commun. Educ.* 50, 283–297. doi: 10.1080/03634520109379256

Titsworth, B. S. (2004). Students' notetaking: the effects of teacher immediacy and clarity. *Commun. Educ.* 53, 305–320. doi: 10.1080/0363452032000305922

Titsworth, B. S., and Kiewra, K. A. (2004). Spoken organizational lecture cues and student notetaking as facilitators of student learning. *Contemp. Educ. Psychol.* 29, 447–461. doi: 10.1016/j.cedpsych.2003.12.001

Uosaki, Y. (2017). Students' Note-taking Styles and Instructed Experiences. *Bull.Faculty Educ. Tamagawa Univ.* 17, 173–185. doi: 10.3390/brainsci8120220

Van Meter, P., Yokoi, L., and Pressley, M. (1994). College students' theory of note-taking derived from their perceptions of note-taking. *J. Educ. Psychol.* 86, 323–338. doi: 10.1037/0022-0663.86.3.323

Wang, C. (2014). An introduction to lecture notes in mathematics. *Spiritual Leaders* 2014:144.

Williams, W. L., Weil, T. M., and Porter, J. C. K. (2012). The relative effects of traditional lectures and guided notes lectures on university student test scores. *Behav. Anal. Today* 13, 12–16. doi: 10.1037/h0100713

Witherby, A. E., and Tauber, S. K. (2019). The current status of students' note-taking: why and how do students take notes?. *J. Appl. Res. Mem. Cogn.* 8, 139–153. doi: 10.1016/j.jarmac.2019.04.002

Yokoi, L. M. (1999). *The Developmental Context of Notetaking: A Qualitative Examination of Secondary-Level Student Notetaking (Junior High School Students, High School Students)*. Ph.D. thesis, Albany: University at Albany, State University of New York.

Zimmerman, B. J. (2008). Investigating Self-Regulation and Motivation: historical Background, Methodological Developments, and Future Prospects. *Am. Educ. Res. J.* 45, 166–183. doi: 10.3102/0002831207312909

Zimmerman, B. J., Bonner, S., and Kovach, R. (1996). *Developing Self-Regulated Learners: beyond Achievement to Self-Efficacy*. Washington DC: American Psychological Association, doi: 10.1037/10213-000

Zimmerman, B. J., and Schunk, D. H. (2011). "Self-Regulated Learning and Performance," in *Handbook of Self-Regulation of Learning and Performance*, eds B. J. Zimmerman and D. H. Schunk (New York: Routledge), 1–12.

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 Empirical Study
PUBLISHED 10 August 2022
DOI 10.3389/fpsyg.2022.936162

Check for updates

OPEN ACCESS

EDITED BY
María Soledad Segretin,
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

REVIEWED BY
Lisa K. Son,
Columbia University, United States
Hye-Jeong Choi,
University of Georgia, United States

*CORRESPONDENCE
Joseph P. Magliano
jmagliano@gsu.edu

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

RECEIVED 04 May 2022
ACCEPTED 20 July 2022
PUBLISHED 10 August 2022

CITATION
Magliano JP, Flynn L, Feller DP, McCarthy KS, McNamara DS and Allen L (2022) Leveraging a multidimensional linguistic analysis of constructed responses produced by college readers.
Front. Psychol. 13:936162.
doi: 10.3389/fpsyg.2022.936162

Leveraging a multidimensional linguistic analysis of constructed responses produced by college readers

Joseph P. Magliano[1]*, Lauren Flynn[2], Daniel P. Feller[1], Kathryn S. McCarthy[1], Danielle S. McNamara[3] and Laura Allen[2]

[1]Department of Learning Sciences, Georgia State University, Atlanta, GA, United States, [2]Department of Educational Psychology, The University of Minnesota, Minneapolis, MN, United States, [3]Department of Psychology, Arizona State University, Tempe, AZ, United States

The goal of this study was to assess the relationships between computational approaches to analyzing constructed responses made during reading and individual differences in the foundational skills of reading in college readers. We also explored if these relationships were consistent across texts and samples collected at different institutions and texts. The study made use of archival data that involved college participants who produced typed constructed responses under thinking aloud instructions reading history and science texts. They also took assessments of vocabulary knowledge and proficiency in comprehension. The protocols were analyzed to assess two different ways to determine their cohesion. One approach involved assessing how readers established connections with themselves (i.e., to other constructed responses they produced). The other approach involved assessing connections between the constructed responses and the texts that were read. Additionally, the comparisons were made by assessing both lexical (i.e., word matching) and semantic (i.e., high dimensional semantic spaces) comparisons. The result showed that both approaches for analyzing cohesion and making the comparisons were correlated with vocabulary knowledge and comprehension proficiency. The implications of the results for theory and practice are discussed.

KEYWORDS

comprehension, college readers, constructed responses, natural language processing, individual differences

10.3389/fpsyg.2022.936162

Introduction

Reading is a fundamental skill for college students. Unfortunately, research suggests that many college students struggle to learn from texts, particularly for the complex expository texts that are often required reading in college classes. It has been estimated that 75% of community college and 50% of four-year students are not adequately prepared for the reading literacy demands of their college coursework (ACT, 2006; Baer et al., 2006; Holschuh and Paulson, 2013). This is in part because the *reading for understanding* that is required for college courses involves a coordination of a variety of processes and strategies that support the construction of an elaborated and coherent mental model for a text (McNamara and Magliano, 2009; Perfetti and Stafura, 2014). Theories of comprehension emphasize that establishing coherence in the mental model is essential for understanding, which means that readers are able to represent how content conveyed in a text is semantically related and how relevant background knowledge is integrated into the mental model (e.g., Kintsch, 1988; Graesser et al., 1994). This representation arises from an interplay of conceptually lower (decoding, lexical access, sentence processes) and higher level processes (inference strategies; Cromley and Azevedo, 2007; Perfetti and Stafura, 2014; Magliano et al., 2020). Foundational skills of reading support processing the printed word, which provides input for the higher order processes that support the construction of a coherent mental model. Some college readers struggle with the foundational skills associated with word reading (Ari, 2016; Halldórsdóttir et al., 2016; Magliano et al., in press; Kopatich et al., 2022), others may be proficient readers, but do not engage in higher order comprehension strategies that support comprehension (Magliano and Millis, 2003; Kopatich et al., 2022), some struggle with both (Magliano et al., in press; Kopatich et al., 2022). These findings suggest that providing support for these different types of struggling college students requires a deeper understanding of the relations between foundational reading skills and comprehension outcomes and better means of evaluating students' different strengths and weaknesses (Magliano and Millis, 2003; Magliano et al., 2011; Perin, 2020).

One approach for studying college readers is through the analysis of constructed responses. Constructed responses produced during reading (e.g., think-alouds) have provided insights into the strengths and challenges of college readers (Magliano and Millis, 2003; Magliano et al., 2011, 2020, in press; Cromley and Wills, 2016; Feller et al., 2020). These responses reveal a variety of strategies that support the construction of a mental model (Pressley and Afflerbach, 1995; Trabasso and Magliano, 1996; McNamara, 2004; Magliano et al., 2011) and are correlated with performance on standardized measures or the foundational skills of reading in college readers (Magliano and Millis, 2003; Magliano et al., 2011, 2020). As such, constructed responses have been useful in the study of underprepared college readers.

Traditionally, constructed responses have been analyzed through expert judgments (i.e., annotation) in which raters use a codebook or scoring rubric to identify different processes and strategies evident in the constructed responses (e.g., Pressley and Afflerbach, 1995; Chi, 1997; McNamara, 2004; Rapp et al., 2007). More recently, there has been a growing interest in leveraging computational analyses to quantify linguistic features of the constructed responses. Natural language processing approaches are sensitive to subtle features of the readers' language that are informative of individual differences in reading proficiency and potentially indicative of sense making (Allen et al., 2016). The present study examines how theoretically motivated and computationally-derived indices of *cohesion* can be used to explore aspects of coherence-building and the extent to which these measures are related to college students' foundational reading skills.

Foundational skills, inference strategies, and comprehension

Reading for understanding involves coordination of a set of skills that support the process of reading and the construction of a coherent mental model (Cromley and Azevedo, 2007; Perfetti and Stafura, 2014; Kopatich et al., 2018). In the context of the present study, we make a distinction between foundational skills and coherence building strategies. Foundational skills involve word, sentence, and discourse level processes that enable readers to construct propositional representations of text sentences that reflect the content that was explicitly conveyed in the text and how propositions are referentially related to one another (Perfetti and Stafura, 2014). Processing words starts the process of activating knowledge needed to build a mental model that is the basis of comprehension, and as such word recognition facilitates processing that supports comprehension (e.g., Perfetti and Stafura, 2014). As such, in the present study we explored the relationship between constructed responses and the foundational skills associated with vocabulary knowledge and discourse comprehension.

Text comprehension involves building a coherent mental model for a text (Graesser et al., 1994). The mental model consists of a network of propositions that reflect content explicitly conveyed and knowledge-based inferences (i.e., inferences generated based on the knowledge that readers activate during reading; Kintsch, 1988). Coherence reflects the relationships that are established between the propositions. That is, coherence building strategies establish how propositions corresponding to sentences are related to one another or integrate relevant background knowledge into the mental model. These strategies range from resolving anaphora (i.e., identifying the references to pronouns) to generating explanatory inferences that specify how propositions are connected (Graesser et al., 1994). Theories of comprehension make a distinction between local and global coherence (McNamara and Magliano, 2009). Specifically, local coherence is established when connections are made between adjacent sentences, whereas global coherence is established when readers establish connections across sentences that are more distally apart. Comprehension requires establishing coherence at both local and global levels (e.g., Graesser et al., 1994).

Contemporary models of reading literacy assume that there are relations between foundational skills of reading and coherence building strategies (Cromley and Azevedo, 2007; Perfetti and Stafura, 2014; Kopatich et al., 2018; Magliano et al., 2020). Some research focuses on testing moderational relationships (e.g., Magliano and Millis, 2003), whereas others focus on testing the presence of mediational relationships (e.g., Cromley and Azevedo, 2007; Kopatich et al., 2018). However, these approaches generally focus on mapping foundational skills to the specific strategies that readers employ during reading (e.g., the frequency that they engage in bridging or paraphrasing). In the present study, we take a different approach by examining the *cohesion* of the constructed responses at multiple levels. Our argument is that cohesion indices can serve as proxies for coherence-building, as they indicate the amount and type of connections that readers are making while reading. Thus, our goal was to examine how these cohesion measures related to individual differences in vocabulary knowledge and comprehension proficiency. Our long-term goal is to leverage these findings for future development of research and interventions directed at using these approaches to better understand the strengths and challenges of college readers.

Importantly, some contemporary frameworks of literacy assume that reading can vary across contexts (Snow, 2002; Britt et al., 2018). Context can involve place and time where the literacy activity takes place, the texts that students read, and the nature of the tasks. The present study therefore takes advantage of archival data that affords the exploration of the extent to which estimated relations between the linguistic features of constructed responses and foundational skills of reading vary are stable across institutions (i.e., community colleges and a university) and texts (i.e., science and history texts).

Analyzing constructed responses

The use of human judgments or expert ratings of constructed responses generated during reading was popularized in the late 1980s and early 1990s (e.g., Ericsson and Simon, 1993; Pressley and Afflerbach, 1995; Trabasso and Magliano, 1996). Since then, it has continued to be a fruitful approach for the study of text comprehension processes (e.g., Rapp et al., 2007; Goldman et al., 2012; Cromley and Wills, 2016). Over the past two decades, advances in natural language processing (NLP) techniques have afforded researchers increased opportunities to examine constructed responses along multiple dimensions (e.g., Magliano et al., 2002, 2011; Magliano and Millis, 2003; McNamara et al., 2006, Millis et al., 2007; Allen et al., 2016). These advances have not only made it easier and more efficient to evaluate constructed responses, but they have also made it possible to calculate linguistic features of the responses that would be difficult for human raters to identify.

Computational analyses of constructed responses are often focused on exploring the extent that readers engage in coherence building strategies (Allen et al., 2016; Magliano and Millis, 2003; Magliano et al., 2011). However, there are several different aims or goals within this body of work. For example, some research uses linguistic analysis to measure the *quality* of the responses (McNamara et al., 2004) while others have aimed to identify the specific *strategies* (e.g., paraphrasing, bridging) that readers are employing (Magliano et al., 2011). Relevant to the current research study, researchers have also examined coherence building by modeling the connections that readers make across various pieces of the text (Allen et al., 2015, 2016), and how the relationship between these various features to comprehension outcomes and individual differences (Allen et al., 2015, 2016). When constructing a coherent mental model of a text, a reader must generate links or connections between concepts. Cohesion is a measure (construct) that captures the ways in which ideas are linked (McNamara et al., 2010). Cohesion analyses involve assessing the degree to which constructed responses overlap with each other. In other words, this approach examines the degree to which readers establish connections across the content within their constructed responses. Recent research suggests that automated analyses of cohesion can be used to assess these connections and can predict individual differences in knowledge and skills as well as learning from the text (Allen et al., 2015, 2016). Cohesion isn't a single thing; it can be assessed in a variety of ways in service of understanding coherence building. Our intent is not to pit different approaches against one another, but rather to more systematically examine how different aspects of cohesion might be used to better understand the coherence-building processes engaged by students, and particularly struggling students.

One aspect of cohesion reflects "what" is being connected (see **Table 1**). Cohesion can be measured using *lexical overlap* – when readers repeat the same content words (nouns, verbs, adverbs, adjectives) from sentence to sentence or echo the words from the text they are reading. Lexical comparisons are made with word matching algorithms that detect if the same words are used in the units of language that are being compared (e.g., Magliano et al., 2011). While researchers may create a dictionary of synonyms to assist these comparisons of constructed responses to the text, others have chosen to constrain lexical comparisons to the exact words in the text (e.g., Magliano et al., 2011). Our approach examines whether words overlap across the constructed responses that readers produced.

For example, consider a student who is reading a text about red blood cells. For the first response, the student states, "Red blood cells are a necessity for the body. They bring carbon dioxide to the cells of the body. The body then turns the oxygen into carbon dioxide. The red blood cells take the carbon dioxide and have it removed." Later in the text, the student states, "Anemia is a condition where not enough oxygen gets into the body. Anemia can make a person feel tired and weak. One time, my doctor told me I was anemic and it made me feel really tired. I guess, anemia must have something to do with problems with your red blood cells. Is blood a cell?" Here, we can see that there are multiple words that the student repeats across constructed responses, such as "body" and "cells." Lexical overlap measures will count the degree to which the reader uses these same words across the responses.

In addition to specific words or close synonyms, another way of assessing cohesion is through *semantic overlap*. Computationally, semantic overlap is calculated through the use of high dimensional semantic spaces (e.g., Latent Semantic Analysis; Landauer and Dumais, 1997; Word2vec; Mikolov et al., 2013). High dimensional semantic spaces are statistical approaches for representing knowledge about words and the world on the basis of a large corpus of texts. Semantic spaces are constructed by analyzing a corpus of thousands or even millions of texts to count how frequently different words co-occur with the assumption that the meaning of words are linked to the words for which they tend to co-occur (Landauer and Dumais, 1997; Burgess et al., 1998; Jones et al., 2006). That is, a set of words share meaning to the extent that occur in similar contexts. For example, "banana" and "apple" share meaning because they are used in similar contexts and thus are closer in semantic space, whereas "banana" and "stapler" are less semantically related. With respect to the example above, a student may not explicitly use the word "body" in their responses; however, if they are talking about their arms or other parts of their body, these responses are still likely to be highly semantically related. The general recommendation is that both lexical and semantic overlap are important for the computational assessment of constructed responses (McNamara et al., 2007; McNamara, 2011; Magliano and Graesser, 2012).

TABLE 1 A summary of the computational approaches for analyzing constructed responses.

Computational approach	Grain size	Lexical	Semantic
Cohesion overlap	Local	Adjacent responses: lexical overlap	Adjacent responses: Word2vec
	Distal	Adjacent (+2) responses: lexical overlap	Adjacent (+2) responses: Word2vec
Source overlap	Source	Proportion of unique words (types) that overlap with source	All responses to source text: Word2vec

One way of evaluating the quality of a reader's mental model and coherence is to evaluate *source overlap* (see **Table 1**). Source overlap reflects the extent to which the reader represents the content words (lexical) or ideas (semantic) that are presented in the text. On the one hand, a reader who is representing too little of the text content may be struggling to identify key ideas or may be off-task. On the other hand, a reader who is representing too much of the text may be including only what is explicit in the text and not making critical inferences to more fully elaborate the mental model. Of course, some words and/or ideas are going to be more important for the construction of an accurate or complete mental model. Thus, one approach to measuring source text overlap is to identify a dictionary of key terms that appear in the source text and calculate how many of those terms appear in the students' responses (e.g., Millis et al., 2007).

In addition to exploring overlap between source and response, researchers have begun to more deeply examine aspects of cohesion *within* a reader's response called *cohesion overlap* (see **Table 1**). Such approaches were initially developed in tools such as Coh-Metrix (Graesser et al., 2004; McNamara et al., 2004) that were designed to calculate indices of cohesion to measure text complexity. These analyses assume that texts that are less cohesive are more complex as they require the reader to generate more inferences in order to maintain coherence (McNamara and Kintsch, 1996). In recent years, researchers have been exploring the extent to which these cohesion indices can be used to infer the coherence of a reader's mental model. For example, Allen et al. (2016) asked readers to generate constructed responses while reading a text about natural selection. NLP techniques were then used to calculate the cohesion of these responses. The results indicated that cohesion values were predictive of text-specific prior knowledge and comprehension. Of relevance to the current study and its focus on coherence-building, measures can be calculated at multiple grain sizes such that you can examine if readers generate connections to the constructed responses that were produced immediately prior as well as ones produced earlier during reading.

Measures of *local overlap* examine how words or ideas are connected across two adjacent sentences in the students' response. The underlying assumption is that local overlap suggests that the reader is engaging in coherence-building processes that are supporting the maintenance of local coherence. While local cohesion is important, it is also critical for readers to be making connections across the larger discourse context. Thus, measures of overlap can be calculated at different grain sizes to examine how readers generate connections across constructed responses produced earlier during reading. The distinction between local and distal can be operationalized in multiple ways. In some cases, local overlap is defined as overlap to only the adjacent sentence and anything beyond that reflects distal or even global overlap (e.g., Magliano et al., 2011). In the current study, we examine two primary grain sizes: overlap for adjacent sentences (local cohesion) in the constructed responses and overlap of two adjacent constructed responses (more distal cohesion). There are numerous grain sizes at which cohesion can be measured; here, we chose two that represent relatively local and distal connections that readers may be making while reading.

Overview of study and research questions

Our overarching aim was to use a multidimensional linguistic analysis of think aloud responses made during reading to examine how students' constructed responses reflect individual differences in foundational skills of reading in college readers. We also explored if these relationships were consistent across texts and samples collected at different institutions. This study uses archival data from Magliano et al. (2020) in which college students produced typed "think aloud" responses while reading expository texts (i.e., science and history texts). This sample contains college readers from two- and four-year institutions wherein some of the students were designated as needing additional support in reading literacy. The participants produced constructed responses while reading a history and a science text. They also were administered an assessment of their proficiencies in the components of foundational reading skills. We conducted the study to address the following two research questions:

> *RQ1: How do readers' foundational reading skills (vocabulary, reading comprehension) relate to the cohesion of constructed responses?*
> *RQ2: To what extent do these relations vary across contexts (i.e., text and sample)?*

Based on prior research, we hypothesized that linguistic analyses of overlap with the text and cohesion of constructed responses would be positively correlated with proficiencies in vocabulary and comprehension (Allen et al., 2015, 2016; Feller et al., 2020). We did not specify *a priori* hypotheses regarding the extent that these relations vary across institutions and texts.

Method

Statement of ethics compliance

The research presented in this article was reviewed by an institutional human subjects compliance board and all participants signed an informed consent form before their participation.

Corpus (data set)

This study is a secondary analysis of data collected in Magliano et al. (2020). The data set includes 560 students from a large, 4-year institution in the Midwest (n = 263), a community college in the Southwest (n = 265), and a community college in the Northeast (n = 32). In the current analysis, we include the 495 students *who completed both measures (described below) and provided demographic information.* Of the participants that reported their gender (n = 392), about 62% identified as female. Of those who selected an age range (n = 373), about 90% were between 18 and 22 years old (range = 18–22 to 50–53). Additionally, of the students that reported whether or not English was their first language (n = 392), approximately 76% reported English as their first language.

Measures

Foundational reading skills

Foundational skills of reading were measured using the Study Aid and Reading Assessment (SARA; Sabatini et al., 2015, 2019). In the current study, we focused on two subtests: vocabulary and reading comprehension. The vocabulary subtest involved presenting target words and participants were asked to select a synonym or topically related word from three choices. The range of possible scores was 0–38. Previously reported reliability estimates range from 0.72 to 0.81 for high school students (Sabatini et al., 2019). In the Reading Comprehension subtest, participants read passages and answered multiple-choice questions about each passage that involved understanding main ideas, locating important details, and drawing inferences. The range of possible scores was 0–22. Previously reported reliability estimates range from 0.80 to 0.85 (Sabatini et al., 2019).

Reading strategy assessment tool

The Reading Strategy Assessment Tool (RSAT; Magliano et al., 2011) was used to collect the think aloud protocols. Participants produced typed constructed responses while reading two texts: Power of Erosion and Louis the XVI and the French Revolution. The Power of Erosion is an earth science text that describes the basic processes underlying erosion and its role in forming geological features. It contains 316 words, 22 sentences, and 5 paragraphs and has a Flesch-Kincaid grade level of 9.9. Participants produced think aloud responses after 7 sentences that were chosen by Magliano et al. (2011) because they afforded bridging and elaborative inference strategies based on a theoretical analysis of the texts. Louis the XVI and the

French Revolution is a history text that makes the case that Louis the XVI has been misunderstood with respect to the events that lead up to the French Revolution. It contains 366 words, 19 sentences, and 4 paragraphs and has a Flesch-Kincaid grade level of 11.7. Participants produced think-aloud responses after 6 sentences.

Data processing

To examine the cohesion of the constructed responses generated by participants, we first aggregated participants' responses for each text and separated them by a paragraph break (Allen et al., 2015). That is, each participant's data was combined into a single science response comprising the seven constructed responses generated while reading that text and a history response (comprising the six constructed responses generated during that text), such that each response was represented as its own paragraph.

We then cleaned the responses by correcting misspellings and editing contractions to be two separate words. NLP tools were then used to calculate the lexical and semantic cohesion of the responses at three different grain sizes, resulting in six cohesion indices (see **Table 1**). For the calculation of all variables except the lexical, source-based indices, we relied on the Tool for the Automatic Analysis of Text Cohesion (TAACO; Crossley et al., 2016), a freely available NLP tool that provides automated measures related to text cohesion. TAACO allows users to calculate indices of local and global cohesion for a given text, ranging from the repetition of words used in the text to the semantic similarity of the paragraphs. For the *lexical, source overlap indices*, we calculated these through simple counts of overlapping lemmas (i.e., *run*, *runs,* and *running* would be considered overlapping because they share the same lemma, *run*) from the source text. For the *local and distal lexical cohesion* indices, we calculated overlap across adjacent responses or adjacent + 2 responses, respectively. For these measures, the overlap was operationalized as the number of words that overlapped with the next (adjacent) or next two (adjacent + 2) responses. These were normalized by the total number of responses that the students produced. These measures provided an indication of lexical cohesion and were intended to represent the times that students were linking ideas and concepts across their constructed responses. For the lexical cohesion indices, the values could be as low as 0 (indicating no lemma overlap) with no explicit cap on the upper end of the range because it is based on the number of lemmas that overlap and thus the length of their responses. In the context of our data, however, the upper end of the range was 20. Here, higher values indicate that more lemmas are repeated across the constructed responses that students generated.

For the *lexical source overlap* measures, we calculated lexical overlap as the proportion of unique words from the text (i.e., the number of word types) that were in the constructed responses to avoid bias from a single keyword in the source text. In other words, we wanted to examine the breadth of topics that students drew on from the source text rather than simply counting the number of times they referenced the primary topic of the text (e.g., erosion). This index was therefore intended to represent the degree to which students were drawing on the concepts from the source text in the responses they generated.

Semantic cohesion was calculated at local (adjacent responses), distal (adjacent + 2 responses), and source text levels. For all three of these indices, semantic cohesion was measured using TAACO (Crossley et al., 2016). TAACO relies on Word2vec to calculate semantic similarity at multiple levels. Similar to LSA (Landauer and Dumais, 1997), Word2vec is an NLP technique for calculating semantic similarity by estimating associations between words using a large corpus of texts (Mikolov et al., 2013). Words in the texts are treated as vectors that are decomposed mathematically and similarity is computed using a cosine value (ranging from 0 to 1). The higher the cosine value, the more strongly two words are semantically related. In comparison to LSA, Word2vec uses of multiple-layered neural network (whereas LSA applies singular value decomposition) and it leverages an increased contextual window that considers words both before and after a target word. After the Word2vec model has been trained, it can detect words that are considered highly similar based on the contexts in which they tend to occur. Therefore, while lexical overlap captures explicit overlap of one's words across responses (i.e., repeating the word "erosion"), semantic overlap affords for less overt word overlap, instead giving way for semantically-similar word overlap (i.e., using words with similar meanings such as "revolution" and "war"). The only difference in the calculation of the three indices related to the segments of text that were being compared: for local and distal indices, we calculated the average Word2vec cosine value between adjacent and adjacent + 2 responses, respectively. For the source overlap index, we calculated similarity with the text that students read.

Given that the students generated constructed responses for two texts, we aggregated the NLP indices at the participant-level for all analyses with the exception of the exploratory genre analysis in research question two.

Data accessibility

The data and R scripts can be accessed online on Open Science Framework (https://osf.io/6fthp/).

Results

Students' aggregated constructed responses contained an average of 85.97 words (SD = 48.11), ranging from 10 to 492. Correlation analyses were conducted to examine relations between readers' proficiencies in foundational skills (i.e., vocabulary and reading comprehension) and linguistic features of constructed responses. Given that variables were not normally distributed, Spearman's Rho was used to examine the relations among variables (Chen and Popovich, 2002). Descriptive statistics and correlations are presented in **Table 1**. Participants on average scored 69.6% (M = 26.46; SD = 6.08) and 55.5% (M = 12.20; SD = 4.29) on the vocabulary and reading comprehension components of SARA, respectively. These scores indicate that, overall, students had relatively strong skills but that these abilities varied considerably across our sample. Performance on these two proficiency components were significantly correlated ($r = 0.63$, $p < 0.001$).

> *RQ1: How do readers' foundational reading skills (vocabulary, reading comprehension) relate to the cohesion of constructed responses?*

To address our first research question, the strength of the relations among vocabulary, reading comprehension, and linguistic features of constructed responses was examined. The correlations among foundational reading skills are presented in **Table 2**.

Word count was a significant positive predictor of vocabulary and reading comprehension scores. This is not surprising, as it is likely that students who are more engaged in reading the texts will produce longer responses on average. Additionally, both vocabulary and reading comprehension scores were significantly, positively related to markers of cohesion associated with word overlap at both local and distal grain sizes. This suggests that participants with higher word knowledge and better comprehension skills were more likely to generate constructed responses that connected to the prior responses they generated.

Similarly, vocabulary and reading comprehension scores were significantly, positively correlated with indices of semantic overlap at both window sizes. These findings are similar to those for lexical overlap and indicate that higher world knowledge and comprehension skills are associated with strong connections at both the word and semantic levels. Measures of semantic overlap differ from measures of word overlap in that they reflect that readers are not only making explicit connections to specific words in their response, but they are also establishing connections by generating responses that refer to similar concepts. Finally, vocabulary and reading comprehension scores were both positively correlated with cohesion indices that reflected connections made to the source text. In particular, overlap at the word and semantic levels were correlated with these scores, a significant, positive correlation was found with type overlap with the source text ($r = 0.28$, $p < 0.001$). This suggests that those with stronger vocabulary skills typically make more explicit lexical connections in their constructed responses than those with weaker vocabulary skills. Finally, reading comprehension scores were positively associated with lemma type overlap with the source text ($r = 0.31$, $p < 0.001$), suggesting that students with higher comprehension scores were more likely to use a greater proportion of lemma types (unique lemmas) from the source text.

A multiple linear regression was used to examine the predictive power of the indices when taken together and controlling for the overall length of the aggregated constructed responses (see **Table 3** for detailed model information). Due to multicollinearity (Variance Inflation Factor >10), one index was not included in the model: Adj. 2-Response Overlap. The overall model significantly predicted vocabulary scores ($R^2 = 0.18$, $F(6,488) = 17.92$, $p < 0.001$). Importantly, while word count was the strongest predictor of vocabulary scores, all of the remaining indices except lexical overlap to the source remained significant. As such, these metrics account for unique variance in vocabulary knowledge even when accounting for the length of the responses. A similar regression examined the ability of these indices to predict reading comprehension scores (see **Table 4** for detailed model information). Due to multicollinearity (Variance Inflation Factor > 10), one index was not included in the model: Adj. 2-Response Overlap. The overall model significantly predicted reading comprehension scores ($R^2 = 0.21$, $F(6,488) = 21.15$, $p < 0.001$). The results of this analysis were similar to those of the vocabulary analyses – namely all indices except lexical overlap to the source text were significant in the model, even after controlling for response length. This indicates that both vocabulary knowledge and reading comprehension are related to cohesion metrics at multiple levels; however, the results also suggest that these variables may not be particularly strong at discriminating amongst these two individual difference measures. As expected, word count was the strongest predictor in the model; however, the majority of the cohesion indices remained significant when this measure was entered in the models. For both models, the source overlap indices were the weakest, indicating that the connections that students established within their constructed responses were more predictive than the connections they made to the source text.

> *RQ2: To what extent do these relations vary across contexts (i.e., text and sample)?*

To address our second research question, we examined the extent to which the relations between cohesion and the SARA scores were stable across text genre (i.e., history, science) and institution (i.e., 2-year, 4-year). Correlations by genre and institution are presented in **Tables 5, 6,** respectively.

We first examined correlations across the two text genres. While the general pattern of results was largely consistent across genres, the strength of these relations varied. Word count had a stronger relation with both vocabulary and reading comprehension in the history text than the science text. Similarly, the relations among vocabulary, reading comprehension, distal response lexical overlap, local response Word2vec, and distal Word2vec were stronger in the history text than the science text. This suggests that constructed responses for the history text used slightly greater word and semantic overlap from the source text than those corresponding to the science text. Overlap with the source text also appeared to vary by text genre, with a stronger negative relationship between lemma type overlap and vocabulary for the history text than the science text. Conversely, the relation between type overlap and vocabulary, though statistically insignificant, was stronger for the science text than the history text.

In terms of institutional differences, there were various relations that varied in magnitude. In general, measures of both word and semantic overlap (i.e., local response lexical overlap, distal response lexical overlap; local response Word2vec, distal response Word2vec) were more strongly correlated with reading comprehension scores in the 2-year institution than the 4-year institution. Additionally, vocabulary scores were more highly correlated with adjacent response overlap for the 4-year institution than the 2-year institution; however, reading comprehension scores were more highly correlated with adjacent response overlap for 2-year institution than the 4-year institution. Lemma overlap also appeared to vary by institution. There was a significant negative correlation between lemma overlap and vocabulary in the 2-year institution but not the 4-year institution. This suggests that the negative relation between

TABLE 2 Overall descriptive statistics and Spearman correlations.

				SARA score correlation	
Index type	**Overlap index**	***M***	**SD**	**Vocabulary**	**Reading comprehension**
Descriptive	Word count	85.97	48.11	0.41***	0.42***
	Adj. response overlap	2.49	1.69	0.26***	0.30***
Lexical cohesion	Adj. 2-response overlap	3.78	2.33	0.28***	0.31***
	Prop. of type overlap	0.10	0.06	0.28***	0.31***
	Adj. response Word2vec	0.78	0.05	0.22***	0.26***
Semantic cohesion	Adj. 2-response Word2vec	0.79	0.06	0.19***	0.20***
	Source similarity Word2vec	0.64	0.15	0.21***	0.21***
	Vocabulary	26.46	6.08	–	0.63***
	Reading comprehension	12.20	4.28	–	–

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

TABLE 3 Linear model predicting SARA vocabulary scores.

Predictor	**Estimate**	***df***	**Std. error**	***t***	***p***	**Partial η^2**
(Intercept)	14.85	1	4.24	3.50	<0.001	
Word count	0.11	1	0.01	7.92	<0.001	0.11
Adj. response overlap	−2.05	1	0.36	−5.70	<0.001	0.06
Prop. of type overlap	−10.32	1	8.71	−1.18	0.237	0.00
Adj. response Word2vec	29.38	1	12.42	2.37	0.018	0.01
Adj. 2-response Word2vec	−24.65	1	9.93	−2.48	0.013	0.01
Source similarity Word2vec	7.81	1	2.60	3.00	0.003	0.02

TABLE 4 Linear model predicting SARA comprehension scores.

Predictor	**Estimate**	***df***	**Std. error**	***t***	***p***	**Partial η^2**
(Intercept)	3.53	1	2.94	1.20	0.23	
Word count	0.07	1	0.01	7.06	<0.001	0.09
Adj. response overlap	−1.16	1	0.25	−4.66	<0.001	0.04
Prop. of type overlap	0.24	1	6.05	0.04	0.968	0
Adj. response Word2vec	29.18	1	8.62	3.39	<0.001	0.02
Adj. 2-response Word2vec	−24.73	1	6.89	−3.59	<0.001	0.03
Source similarity Word2vec	3.99	1	1.81	2.21	0.028	0.01

TABLE 5 Descriptive statistics and Spearman correlations by genre.

				SARA score correlation	
Index type	**Overlap index**	***M***	**SD**	**Vocabulary**	**Reading comprehension**
				History	
Descriptive	Word count	83.77	48.11	0.41***	0.44***
	Adj. response overlap	2.61	1.99	0.26***	0.31***
Lexical cohesion	Adj. 2-response overlap	3.92	2.70	0.28***	0.31***
	Prop. of type overlap	0.08	0.05	0.23***	0.29***
	Adj. response Word2vec	0.77	0.08	0.19***	0.22***
Semantic cohesion	Adj. 2-response Word2vec	0.77	0.09	0.16***	0.17***
	Source similarity Word2vec	0.56	0.18	0.16***	0.17***
				Science	
Descriptive	Word count	88.17	48.05	0.37***	0.36***
	Adj. response overlap	2.37	1.72	0.23***	0.24***
Lexical cohesion	Adj. 2-response overlap	3.64	2.38	0.24***	0.24***
	Prop. of type overlap	0.12	0.06	0.28***	0.28***
	Adj. response Word2vec	0.79	0.06	0.16**	0.17**
Semantic cohesion	Adj. 2-response Word2vec	0.81	0.07	0.15**	0.13**
	Source similarity Word2vec	0.72	0.17	0.23***	0.23***

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

vocabulary and lemma overlap was stronger for participants in the 2-year institution. Lemma type (i.e., unique lemmas) also varied by institution, with a stronger relation between vocabulary and lemma type in the 4-year institution than the 2-year institution.

Discussion

The purpose of the present study was to explore the relations between college students' foundational reading skills and their coherence-building processes. We used natural language processing tools to calculate a multidimensional analysis coherence-building strategies based on indices reflecting cohesion. More specifically, we examined cohesion at both lexical and semantic levels and across multiple levels of connection (i.e., local overlap, global overlap, source text overlap). We addressed two research questions. The first research question (How do readers' foundational reading skills (vocabulary, reading comprehension) relate to the cohesion of constructed responses?) pertained to whether the different approaches were correlated with the foundational skills of vocabulary knowledge and comprehension proficiency. The results of the present study are consistent with prior research that shows that the computational analysis of constructed response associated with coherence building strategies are correlated with individual differences in foundational skills of reading (Magliano and Millis, 2003; Allen et al., 2015, 2016; Feller et al., 2020; Magliano et al., 2020). All six indices of cohesion were positively and significantly correlated with vocabulary and reading comprehension proficiency suggesting that readers with stronger foundational skills in reading were creating more connections as they read. The strongest relations were with the explicit lexical connections made across the larger response context and to the source text. Relations between the cohesion indices and reading skill were also somewhat stronger for reading comprehension as compared to vocabulary, potentially pointing to the importance of specific comprehension skills that elicit coherence-building amongst readers.

With respect to the second research question (To what extent do these relations vary across contexts (i.e., text and sample)?), we examined how the relations described above varied as a function of text genre (science, history) and across institutional contexts (community college, University). Our findings suggest that the relations between foundational reading skill and response cohesion were relatively stable across these contexts with some differences in magnitude. For students at a 2-year community college, cohesion measures were more strongly related to proficiency in reading skill than to vocabulary. By contrast, the students enrolled at the four-year institution showed an inverse effect, such that cohesion measures were more strongly correlated with vocabulary score as compared to reading comprehension proficiency. Although we caution over interpreting these modest effects, they may reflect the fact that community college and universities have different admissions criteria. Community colleges are open access and accept anyone who applies, whereas universities typically have admissions criteria based on performance on standardized tests (e.g., ACT,

TABLE 6 Descriptive statistics and Spearman correlations by institution.

Cohesion type	Overlap index	*M*	SD	SARA score correlation: Vocabulary	SARA score correlation: Reading comprehension
				2-year institution	
Descriptive	Word count	85.99	49.50	0.39***	0.43***
	Adj. response overlap	2.59	1.84	0.24***	0.35***
Lexical cohesion	Adj. 2-response overlap	3.87	2.49	0.28***	0.34***
	Prop. of type overlap	0.11	0.06	0.24***	0.33***
	Adj. response Word2vec	0.78	0.05	0.22***	0.33***
Semantic cohesion	Adj. 2-response Word2vec	0.79	0.05	0.200**	0.26***
	Source similarity Word2vec	0.65	0.16	0.18**	0.32***
				4-year institution	
Descriptive	Word count	85.95	41.76	0.42***	0.41***
	Adj. response overlap	2.40	1.53	0.31***	0.27***
Lexical cohesion	Adj. 2-response overlap	3.69	2.18	0.31***	0.29***
	Prop. of type overlap	0.10	0.05	0.39***	0.32***
	Adj. response Word2vec	0.78	0.06	0.26***	0.22***
Semantic cohesion	Adj. 2-response Word2vec	0.78	0.07	0.22***	0.16*
	Source similarity Word2vec	0.62	0.15	0.32***	0.16**

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

SAT). As such, there will likely be differences in the distribution of foundational skills of reading given that lower skilled readers will be admitted to community college than at universities. As such, there could be greater variability in foundational skills of reading in a community college setting, which underscores the challenges of providing specific support for college students who may be underprepared to read for their college coursework.

With respect to the stability across texts, we found similar relationships between the history and science texts used in this study. While there are disciplinary skills for comprehending and using history and science tests (Shanahan and Shanahan, 2008), there are also likely a common set of skills that support coherence building strategies. That is, readers need to accurately process words, activate relevant word and general knowledge, and use that knowledge to establish connections across content (Perfetti and Stafura, 2014). As such, variability in performance on these standardized tests are consistently correlated with computational indices of coherence building strategies. Given that there are only two texts in this data set, it is difficult to discern if the small differences across correlations reflect idiosyncratic variations that occur from text to text or if these differences point to more generalizable effects that might emerge in students reading of different genres or disciplines.

Implications

These results have important theoretical implications. Theories of text comprehension emphasize the importance of coherence building strategies for comprehension but are arguably agnostic about the implications of individual differences in the foundational skills of reading on those strategies (McNamara and Magliano, 2009). On the other hand, theories of reading typically explicitly specify the relationships between foundational skills of reading and comprehension proficiency (e.g., Gough and Tunmer, 1986; Cromley and Azevedo, 2007). The results of the present study indicate that theoretical frameworks that describe how foundational skills of reading and coherence building strategies are coordinated are warranted (Perfetti and Stafura, 2014; Kopatich et al., 2018; Magliano et al., 2020). The results of the present study suggest that the computational analysis of constructed responses could be a valuable tool to test and refine these frameworks.

With respect to implications for practice, this study was motivated by the challenge of finding ways to learn about and help support early college students read for their coursework. This goal is motivated by the fact that many students come to college not ready to read (ACT, 2006; Baer et al., 2006; Holschuh and Paulson, 2013). While the computational analysis of constructed responses can be used to assess individual differences (e.g., Magliano et al., 2011), they are often used in computer-based interventions as a way to provide feedback to students (e.g., McNamara et al., 2004). Both the text comparison and the constructed response cohesion approaches assess the extent that readers are making connections between their thinking about the text and how their thoughts are grounded in the text. The computational approaches that we have used have been limited to assessing the global quality of constructed responses (e.g., McNamara et al., 2004). The results of this study suggest that computational approaches to examining constructed responses may be able to drive more specific feedback about students' coherence-building processes when reading. For example, feedback could be given regarding the extent that students are grounding the text in the discourse context and establishing local and global coherence. The connections that students make could be visualized such that they can more easily see the connections they have made to the prior text content. Moreover, such analyses of constructed responses may help to reveal different profiles of strengths and challenges of readers (Rapp et al., 2007; McMaster et al., 2012; Kopatich et al., 2022). It is possible that some struggling college readers do not connect their thoughts to the text, whereas others show evidence of establishing local coherence, but need additional support learning strategies to establish global coherence. The present study was not designed to determine exactly how the indices explored in this study could be used in such applications, but it demonstrates that such research is warranted.

Limitations and future directions

One limitation of this study is that we relied on a secondary data analysis and focused upon limited measures of foundational reading skill. Vocabulary and reading comprehension are but a few foundational reading skills that are thought to impact reading processes and outcomes. In the same vein, we selected a small set of theoretically-motivated linguistic indices related to dimensions of cohesion. These indices were selected to specifically target our dimensions of interest, but do not reflect all possible ways of capturing the cohesion of language. For example, the Tool for the Automatic Assessment of Cohesion (Crossley et al., 2016) provides more than 150 indices related to cohesion. Further, comprehension processes and individual differences are likely to manifest in other dimensions of language (e.g., lexical features, syntactic complexity) that were noted explored in this study. As such, it is important to acknowledge that the indices used in this study did not account for a large proportion of the variance in students' vocabulary knowledge or comprehension proficiency. Our intent was not to fully model all of the proficiencies and processes related to reading comprehension (e.g., SEM) nor to produce an algorithm that can accurately predict readers' individual differences, but rather to conduct a theoretically-driven exploratory analysis of a particular aspect of coherence-building (i.e., cohesion) as it might relate to foundational skills of reading.

However, the current work gives support to the need for more large-scale studies that "put together the pieces" of comprehension. As is core to the assumption of this work, reading involves complex, interactive processes that operate on multiple levels (Perfetti and Stafura, 2014). Thus, future work should include additional measures of foundational reading that can provide a more stable representation of the construct(s) in question (e.g., Cromley and Azevedo, 2007; Kopatich et al., 2018; Magliano et al., 2020) as well as additional relevant individual differences that are known to impact comprehension processes (e.g., prior knowledge; McCarthy and McNamara, 2021). Similarly, future work should extend the type and quantity of linguistic indices. Such studies would both increase our ability to model student learning through more accurate algorithms and through deeper understanding of the relations between student constructed responses and the underlying comprehension processes.

Additionally, there are significant challenges to support underprepared college readers (Calcagno and Long, 2008; Crisp and Delgado, 2014; Bailey et al., 2016; Ganga et al., 2018), finding ways to support college readers is of critical importance (Boylan and Trawick, 2015). There are dramatic changes to how students are provided support (Cormier and Bickerstaff, 2019). Traditionally, institutionally designated underprepared students have been required to pass literacy courses before they can take credit bearing courses (e.g., Bailey et al., 2016), and indeed a large portion of the sample in the present study were assigned to such a course. However, new approaches involve additional support while students take credit bearing courses. Tutoring systems could be developed specifically for struggling college readers that utilize constructed responses to provide individualized training to refine coherence building strategies. We believe the indices used in this study could be incorporated into such systems. However, more research is needed to develop these systems and to refine the computational algorithm that would be used to provide feedback.

A final issue to consider involves the manner in which the think aloud protocols were produced. Specifically, RSAT has participants type their thoughts, whereas thinking aloud typically involves orally producing thoughts (e.g., Ericsson and Simon, 1993). Prior research has shown that typed and orally produced think aloud protocols for college students are very similar in terms of the presence of comprehension strategies as identified by human judges (Muñoz et al., 2006). However, it is an open question as to whether the relationship between measures of coherence and comprehension outcomes and individual differences would be different if protocols were produced by typing or orally producing them. Speech to text software was not a viable data collection option when RSAT was created, whereas it is now. As such, it is important to assess how these measures might be affected by the modality in which think aloud protocols are produced.

Conclusion

Our results indicate that the computational analysis of constructed responses produced by college readers has promise for theory and applications. Learning how to support underprepared college readers requires more research in the cognition of reading for understanding (Magliano et al., 2020; Perin, 2020). Constructed responses have promise in the context of this research. Moreover, they have promise in the context of intelligent tutoring systems that teach strategies to support reading for understanding (Graesser and McNamara, 2011). In this study we identified new approaches to analyze constructed responses produced during reading that will have important implications on the endeavors.

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.

Ethics statement

The studies involving human participants were reviewed and approved by University Research Services and Administration at Georgia State University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JM was the lead investigator on the study, contributed to the conceptualization of the study, and was the lead author. LF was a co-investigator on the project, contributed to the conceptualization of the study and conducting the analyses, and was responsible for writing the results sections of the manuscript. DF was a co-investigator on the project and contributed to the conceptualization of the study and the writing of the manuscript. LA was a co-investigator on the project, contributed to the conceptualization of the study and was responsible for overseeing the analysis and the writing of the results section. KM was a co-investigator on the study, contributed to the conceptualization of the study and assisted in the writing of the manuscript. DM was a co-investigator on the study, contributed to the conceptualization of the study and assisted in the writing of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by grants awarded by the Institute of Education Sciences, US Department of Education (R305A190063 and R305A150193).

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.

Author disclaimer

All opinions expressed in this manuscript are those of the authors and not the funding agencies. None of the authors have a financial stake in the publication of this research and therefore there are no perceived conflicts of interest.

References

ACT (2006). *Reading between the lines: What the ACT reveals about college readiness in reading.* Iowa City, IA: ACT.

Allen, L. K., Jacovina, M. E., and McNamara, D. S. (2016). "Cohesive features of deep text comprehension processes," in *Proceedings of the 38th Annual Meeting of the cognitive science society in Philadelphia, PA*, eds J. Trueswell, A. Papafragou, D. Grodner, and D. Mirman (Austin, TX: Cognitive Science Society), 2681–2686. doi: 10.1044/2021_LSHSS-20-00117

Allen, L. K., Snow, E. L., and McNamara, D. S. (2015). "Are you reading my mind?: Modeling students' reading comprehension skills with natural language processing techniques," in *Proceedings of the fifth international conference on learning analytics and knowledge*, (New York: ACM), 246–254. doi: 10.1145/2723576.2723617

Ari, O. (2016). Word recognition processes in college-age students' reading comprehension achievement. *Commun. College J. Res. Pract.* 40, 718–723. doi: 10.1080/10668926.2015.1098578

Baer, J. D., Cook, A. L., and Baldi, S. (2006). *The Literacy of America's College Students*. Virginia: American Institutes for Research.doi: 10.1037/e539672012-001

Bailey, T., Bashford, J., Boatman, A., Squires, J., Weiss, M., Doyle, W., et al. (2016). *Strategies for postsecondary students in developmental education: A practice guide for college and university administrators, advisors, and faculty*. Washington, DC: Institute of Education Sciences, What Works Clearinghouse.

Boylan, H. R., and Trawick, A. R. (2015). Contemporary developmental education: Maybe it's not as bad as it looks. *Res. Teach. Dev. Educ.* 31, 26–37.

Britt, M. A., Rouet, J.-F., and Durik, A. M. (2018). *Literacy beyond text comprehension: A theory of purposeful reading*. New York, NY: Routledge.doi: 10.4324/9781315682860

Burgess, C., Livesay, K., and Lund, K. (1998). Explorations in context space: Words, sentences, discourse. *Discours. Process.* 25, 211–257. doi: 10.1080/01638539809545027

Calcagno, J. C., and Long, B. T. (2008). *The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance (No. w14194)*. Cambridge: National Bureau of Economic Research. 10.3386/w14194

Chen, P. Y., and Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures. (Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07–139)*. Newbury Park, CA: Sage.doi: 10.4135/9781412983808

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *J. Learn. Sci.* 6, 271–315. doi: 10.1207/s15327809jls0603_1

Cormier, M., and Bickerstaff, S. (2019). "Research on developmental education instruction for adult literacy learners," in *The wiley handbook of adult literacy*, ed. D. Perin (Hoboken, NJ: John Wiley & Sons, Inc), 541–561. doi: 10.1002/9781119261407.ch25

Crisp, G., and Delgado, C. (2014). The impact of developmental education on community college persistence and vertical transfer. *Commun. College Rev.* 42, 99–117. doi: 10.1177/0091552113516488

Cromley, J. G., and Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *J. Educ. Psychol.* 99, 311–325. doi: 10.1037/0022-0663.99.2.311

Cromley, J. G., and Wills, T. W. (2016). Flexible strategy use by students who learn much versus little from text: Transitions within think-aloud protocols. *J. Res. Read.* 39, 50–71. doi: 10.1111/1467-9817.12026

Crossley, S. A., Kyle, K., and McNamara, D. S. (2016). The Tool for the Automatic Analysis of Text Cohesion (TAACO): Automatic Assessment of Local, Global, and Text Cohesion. *Behav. Res. Methods* 48, 1227–1237. doi: 10.3758/s13428-015-0651-7

Ericsson, K. A., and Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (revised ed.)*. Cambridge: MIT Press.doi: 10.7551/mitpress/5657.001.0001

Feller, D. P., Magliano, J. P., O'Reilly, T., Sabatini, J., and Kopatich, R. D. (2020). Relations between Component Reading Skills, Inferences, and Comprehension Performance in Community College Readers. *Discour. Process.* 57, 473–490. doi: 10.1080/0163853X.2020.1759175

Ganga, E., Mazzariello, A., and Edgecombe, N. (2018). *Developmental education: An introduction for policymakers*. Denver, CO: Education Commission of the States.

Goldman, S. R., Braasch, J. L., Wiley, J., Graesser, A. C., and Brodowinska, K. (2012). Comprehending and learning from Internet sources: Processing patterns of better and poorer learners. *Read. Res. Quart.* 47, 356–381.

Gough, P. B., and Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remed. Special Educ.* 7, 6–10. doi: 10.1177/074193258600700104

Graesser, A. C., and McNamara, D. S. (2011). Computational of multilevel discourse comprehension. *Topics Cogn. Sci.* 3, 371–398. doi: 10.1111/j.1756-8765.2010.01081.x

Graesser, A. C., McNamara, D. S., Louwerse, M. M., and Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behav. Res. Methods Instrum. Comput.* 36, 193–202. doi: 10.3758/BF03195564

Graesser, A. C., Singer, M., and Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychol. Rev.* 101, 371–395. doi: 10.1037/0033-295X.101.3.371

Halldórsdóttir, L. Y., Valdimarsdottir, H., and Arntzen, E. (2016). Effects of a multicomponent intervention to improve reading fluency in adult students. *Behav. Intervent.* 32, 3–20. doi: 10.1002/bin.1452

Holschuh, J. P., and Paulson, E. J. (2013). *The terrain of college developmental reading*. Available online at: https://www.crla.net/docs/TheTerrainofCollege91913.pdf (accessed July 2013).

Jones, M. N., Kintsch, W., and Mewhort, D. J. (2006). High-dimensional semantic space accounts of priming. *J. Mem. Lang.* 55, 534–552. doi: 10.1016/j.jml.2006.07.003

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychol. Rev.* 95, 163–182. doi: 10.1037/0033-295X.95.2.163

Kopatich, R. D., Magliano, J. P., Millis, K. K., Parker, C. P., and Ray, M. (2018). Understanding how language-specific and domain-general resources support comprehension. *Discour. Process.* 56, 530–552. doi: 10.1080/0163853X.2018.1519358

Kopatich, R. D., Magliano, J. P., and Santuzi, A. (2022). *Adult reader profiles by reading proficiency and inference processes used during reading*. Ph.D.thesis.

Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato'sp roblem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240. doi: 10.1037/0033-295X.104.2.211

Magliano, J. P., and Graesser, A. C. (2012). Computer-based assessment of student-constructed responses. *Behav. Res. Methods* 44, 608–621. doi: 10.3758/s13428-012-0211-3

Magliano, J. P., Higgs, K., Santuzzi, A., Tonks, S. M., O'Reilly, T., Sabatini, J., et al. (2020). Testing the Inference Mediation Hypothesis in a Post-Secondary Context. *Contemp. Educ. Psychol.* 61:101867. doi: 10.1016/j.cedpsych.2020.101867

Magliano, J. P., and Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cogn. Instruct.* 21, 251–283. doi: 10.1207/S1532690XCI2103_02

Magliano, J. P., Millis, K. K., Levinstein, I., Boonthum, C., and Rsat Team. (2011). Assessing comprehension during reading with the Reading Strategy Assessment Tool (RSAT). *Metacogn. Learn.* 6, 131–154. doi: 10.1007/s11409-010-9064-2

Magliano, J. P., Talwar, A., Feller, D. P., Wang, Z., O'Reilly, T., and Sabatini, J. (in press). Exploring thresholds in the foundational skills for reading and comprehension outcomes in the context of postsecondary readers. *J. Learn. Disord.* doi: 10.1177/00222194221087387

Magliano, J. P., Wiemer-Hastings, K., Millis, K. K., Muñoz, B. D., and Mcnamara, D. (2002). Using latent semantic analysis to assess reader strategies. *Behav. Res. Methods Instru. Comp.* 34, 181–188. doi: 10.3758/bf03195441

McCarthy, K. S., and McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework. *Educ. Psychol.* 56, 196–214. doi: 10.1080/00461520.2021.1872379

McMaster, K. L., Van den Broek, P., Espin, C. A., White, M. J., Rapp, D. N., Kendeou, P., et al. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learn. Individ. Differ.* 22, 100–111. doi: 10.1016/j.lindif.2011.11.017

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discour. Process.* 38, 1–30. doi: 10.1207/s15326950dp3801_1

McNamara, D. S. (2011). Computational methods to extract meaning from text and advance theories of human cognition. *Topics Cogn. Sci.* 3, 3–17. doi: 10.1111/j.1756-8765.2010.01117.x

McNamara, D. S., Boonthum, C., Levinstein, I. B., and Millis, K. (2007). "Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms," in *Handbook of latent semantic analysis*, eds T. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch (Mahwah, NJ: Erlbaum), 227–241.

McNamara, D. S., and Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discour. Process.* 22, 247–288. doi: 10.1080/01638539609544975

McNamara, D. S., Levinstein, I. B., and Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behav. Res. Methods Instru. Comp.* 36, 222–233. doi: 10.3758/BF03195567

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., and Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discour. Process.* 47, 292–330. doi: 10.1080/01638530902959943

McNamara, D. S., and Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychol. Learn. Motivat.* 51, 297–384. doi: 10.1016/S0079-7421(09)51009-2

McNamara, D. S., O'Reilly, T. P., Best, R. M., and Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *J. Educ. Comp. Res.* 34, 147–171. doi: 10.2190/1RU5-HDTJ-A5C8-JVWE

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Adv. Neural Inform. Process. Syst.* 26, 3111–3119.

Millis, K., Magliano, J., Wiemer-Hastings, K., Todaro, S., and McNamara, D. S. (2007). "Assessing and improving comprehension with latent semantic analysis," in *Handbook of latent semantic analysis*, eds T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch (Mahwah: Lawrence Erlbaum Associates Publishers), 207–225.

Muñoz, B., Magliano, J. P., Sheridan, R., and McNamara, D. S. (2006). Typing versus thinking aloud when reading: Implications for computer-based assessment and training tools. *Behav. Res. Methods Instru. Comp.* 38, 211–217. doi: 10.3758/bf03192771

Perfetti, C., and Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Sci. Stud. Read.* 18, 22–37. doi: 10.1080/10888438.2013.827687

Perin, D. (2020). "Reading, Writing, and Self-Efficacy of Low-Skilled Postsecondary Students," in *The wiley handbook of adult literacy*, ed. D. Perin (Blackwell: Wiley), 237–260. doi: 10.1002/9781119261407.ch11

Pressley, M., and Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.doi: 10.2307/358808

Rapp, D. N., Broek, P. V. D., McMaster, K. L., Kendeou, P., and Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Sci. Stud. Read.* 11, 289–312. doi: 10.1080/10888430701530417

Sabatini, J., Bruce, K., Steinberg, J., and Weeks, J. (2015). SARA reading components tests, RISE forms: Technical adequacy and test design. *ETS Res. Rep. Ser.* 2015, 1–20.doi: 10.1002/ets2.12076

Sabatini, J., Weeks, J., O'Reilly, T., Bruce, K., Steinberg, J., and Chao, S.-F. (2019). SARA Reading Components Tests, RISE Forms: Technical Adequacy and Test Design. *ETS Res. Rep. Ser.* 2019, 1–30. doi: 10.1002/ets2.12269

Shanahan, T., and Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educ. Rev.* 78, 40–59. doi: 10.17763/haer.78.1.v62444321p602101

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.

Trabasso, T., and Magliano, J. P. (1996). Conscious understanding during comprehension. *Discour. Process.* 21, 255–287. doi: 10.1080/01638539609544959

TYPE Systematic Review
PUBLISHED 06 September 2022
DOI 10.3389/fpsyg.2022.980133

Check for updates

OPEN ACCESS

EDITED BY
María Soledad Segretin,
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

REVIEWED BY
Gonzalo Daniel Sad,
CONICET French-Argentine International Center for Information and Systems Sciences (CIFASIS), Argentina
Syu Siang Wang,
Academia Sinica, Taiwan

*CORRESPONDENCE
Tan Shaojie
787788545@qq.com

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

RECEIVED 28 June 2022
ACCEPTED 16 August 2022
PUBLISHED 06 September 2022

CITATION
Shaojie T, Samad AA and Ismail L (2022) Systematic literature review on audio-visual multimodal input in listening comprehension.
Front. Psychol. 13:980133.
doi: 10.3389/fpsyg.2022.980133

Systematic literature review on audio-visual multimodal input in listening comprehension

Tan Shaojie[1]*, Arshad Abd Samad[2] and Lilliati Ismail[3]

[1]School of English, Anhui International Studies University, Anhui, China, [2]School of Education, Taylor's University, Selangor, Malaysia, [3]Faculty of Educational Studies, Universiti Putra Malaysia, Selangor, Malaysia

The purpose of this study is to discuss the effects of audiovisual input on second language acquisition (SLA) and the factors that influence the difficulty of audiovisual learning through a systematic literature review. Prior to this systematic review, in this paper, we searched papers on related topics for the past 10 years from 2012 to 2022, and found 46 journal papers that met the research criteria. They can basically represent the scholarly work in this field. The 46 studies were published in journals indexed in Google Scholar, Eric, Scopus, and Wiley Library. Databases were selected according to a set of inclusion and exclusion criteria. The following conclusions are drawn from the literature review: Audiovisual input can provide more authentic language input and more adequate and richer multimodal cultural and situational contexts, which can better promote learners' understanding of the content and stimulate learners' interest in participating in listening comprehension tasks. The influencing factors of multimodal input on listening difficulty include subtitles, video type, and the relationship between the audio and visual input.

KEYWORDS

systematic literature review (SLR), audiovisual input, multimodality, second language (L2) acquisition, listening comprehension

Introduction

With the development of science and technology, second language (L2) listening teaching is changing from traditional audio teaching to audio-visual teaching, and academic circles have become more and more concerned with related research in the use of audio-visuals (Zhyrun, 2016; Namaziandost and Nasri, 2019; Arbab, 2020). Audio-visual input activates both visual and auditory perceptions while audio input only activates auditory perception (Surguladze et al., 2001; Campbell, 2008) and hence, audio-visual input can be considered as a kind of multi-modal input, which is mainly manifested in the form of image (dynamic), sound, and subtitles, and embodies three meta-functions of image, text, and action. Relevant studies have shown that the efficiency of obtaining information through the combination of audio and visuals is far more effective than through either one of the inputs on its own, and that the information is more durable in the memory (Chao et al., 2015; He et al., 2015). Obviously, the distinction between just listening and listening comprehension is important as listening can be an ability to listen without any interpretation and response while listening comprehension involves the complex process of the brain's selection and processing of information. In this complex listening comprehension process, there are some external factors that interfere in or promote listening comprehension, such as the use of the two different input forms of audio and video.

The importance of examining the effects of audio and visual input, both as a single input and when combined, is due to the increased use of modern technology in the classroom and for varied educational purposes. With the aid of multimedia technology, some large-scale tests, such as TOEFL's iBT including national based examinations such as China's CET-4 and CET-6 internet-based tests, began to use audiovisual input materials such as pictures or videos (Wang et al., 2014). As early as the 1990s, a large number of multimedia materials were used in second/foreign language classrooms, and researchers began to explore the impact of visual teaching materials on learning and learners' psychological cognitive factors. Since then, there has been increasing interest in the use of audiovisual input in listening tests. Research related to this has developed for more than 20 years. However, there is still no conclusive among the findings, especially with respect to how audio-visual input affects second language learning and what elements of this input educators should pay attention to.

In exploring the effect of video (audio and visuals combined) and audio-only on listening comprehension, scholars have made further discoveries through empirical research. Some scholars believe that compared with audio-only materials of the same content, audio and visual combined materials can reduce the difficulty of listening (Seeber et al., 2010; Zhyrun, 2016). Ginther (2002) also found that the use of videos can complement audio information with scene context. Hu and Zhang (2013) found through empirical research involving Chinese speaking students that the multimodal combination of audio and video with English subtitles has the greatest effect on promoting students' listening content comprehension, followed by audio and video with Chinese subtitles, and audio and video without subtitles, while audio alone has the least effect. Although the video will distract students' attention to a certain extent and cause a "split" effect, with the cooperation of the target language, it can resist the interference of the video to a certain extent, which is beneficial to students' listening comprehension (Cohen, 2014). Some differing voices were also found in the study which showed that excessive or irrelevant or mismatched audio-visual information may interfere with the audio-visual comprehension process (Canning-Wilson and Wallace, 2000).

The theoretical basis of L2 audio-visual multimodal input mainly includes the Input Hypothesis (Gregg and Krashen, 1986) and Cognitive Load theory (Sweller, 2010). Based on these two theories, this study attempts to retrieve and sort out the related research results of L2 audio-visual multimodal input, and then make a systematic literature review on this basis focusing on audio-visual materials in listening comprehension. As will be discussed in the next section, much research on audio-visual input in language learning conducted prior to the last decade have focused on two central concerns. The first is research on the influence of audio-visual multimodal input on second language acquisition, and the other is the research on the factors that affect the difficulty of second language audio-visual multimodal input.

From the perspective of conducting a systematic literature review, the research questions play a critical role in determining the search strategy, data extraction, and analysis. The research questions identified in this study are given below:

1. What are the effects of audiovisual multimodal input on second language listening comprehension?
2. What are the key factors that affect listening comprehension performance when using visual input?

This paper starts with the introduction to the study, followed by the influencing factors of audio-visual multimodal input difficulty, then, the research protocol and the execution of the systematic literature review are described. This is followed by the findings and discussion of this study. Finally, the conclusion of the study is presented.

Influencing factors of audio-visual multimodal input difficulty

Based on the above research results, since the 1990s, researchers have carried out empirical studies to examine the factors affecting the difficulty of multimodal input of second language audio-visual, among which three factors have received more attention: audiovisual input, text type, personal factors (Bloomfield et al., 2010; Peters and Muñoz, 2020).

Visual input

Existing listening tests can use five different information input methods—audio-only input and four visual inputs, namely context-only still images, context-only video, content still images, and content video (Ockey, 2007). Ockey proposed that with the different input methods of information in the listening test, the way test-takers process information will also change, which will lead to differences in test performance and thus affect the construct validity of the listening test. Therefore, most of the existing research focuses on comparing the impact of different audiovisual input methods on test scores and the performance of test-takers (Ockey, 2007). Earlier studies mainly compared the effect of audio-only input and video recording.

Among the studies that examine the effect of different audiovisual input methods, Rajabi et al. (2021) found in a study involving 91 second language learners that there was no significant difference in student achievement between audio-only and video-mediated exams. He also found that some students, apparently distracted by visual input, chose not to look at the screen.

Coniam (2001) compared the difference between audio-only input and video input with 104 Hong Kong English learners using open-ended test questions. The results showed that the audio-only group performed better than the video group, but the difference was not statistically significant. Moreover, the subjects in the video group did not think that using video as a medium in listening was helpful for listening comprehension, nor did the audio group think that using audio was more beneficial. Conversely, 36% of test-takers reported not looking at the screen at all during the test, and a small number of test-takers found the video to be distracting. Cubilo and Winke (2013) used writing and Note-taking tasks to measure listening comprehension and found that the quality of writing after listening was the same under the conditions of visual and auditory input, but the subjects' note-taking behavior was different—the quality of note-taking involving visual input decreased significantly. In contrast to these findings, however, Wagner (2010) found that the video group performed 6.5% higher on the post-test than the audio group, and the difference was significant. He believes that the reason is that the non-verbal information in the video helps the subjects to improve their performance.

There are also studies comparing the effect of different visual input modalities. Ockey (2007), for example, compared the different performances of the subjects when the listening test used a series of still images that provided only the context and only videos of the context. He observed six college students whose native language was not English, and collected data by means of retrospective reports, interviews, and video recordings, and found that in these two different input presentation methods, the subjects' involvement in visual input was manipulated as the time when the subject's eyes were in contact with the display screen. Under the still-picture condition, the subjects had little involvement with the visual input and responded consistently. Most of the subjects believed that still images were only useful in the initial context of listening comprehension and did not help much afterward, but also did not interfere. However, under the video recording condition, there were strong individual differences in the way participants were involved in visual information. Some subjects thought video recording was very helpful for listening comprehension, while others thought video recording was very disruptive to listening comprehension.

Additionally, studies have explored learners' preference for visual input and its relationship to test performance. Cheng and Chau (2016) used a questionnaire to examine Japanese English learners' attitudes toward video-based listening tests and found that 91.9% of students preferred video-based listening tests compared to audio-only tests. However, the study did not answer whether students who prefer video-based tests benefit from visual aids and achieve better test scores.

Suvorov's (2009) research revolves around this problem but has not found a conclusive answer. Overall, test takers had different preferences for different input methods, but their preference for a particular input method did not significantly improve test scores. For example, some candidates are more comfortable with video input than audio-only input, but their performance on the video-input part of the listening test is not necessarily better than the audio-input part. Interestingly, however, students who preferred video input scored significantly higher on the audio listening section than the video section.

Text type

The second factor influencing the difficulty of audiovisual multimodal input is text type. Ginther (2002), for example, compared the effects of different types of visual input on TOEFL listening comprehension and found that there was an interaction between text types and visual input types. In his research, Ginther combined visual input with text types. He used context-only visual input (a still photo with the speaker and scene) for two-person conversations, short conversations, and academic discussion sections, and a series of still photos and content visual input for mini-talk sections, including photographs, diagrams, and/or diagrams related to listening input. He found that the content visual input accompanying small dialogues and the situational visual input accompanying academic discussions were helpful for listening comprehension, while the situational visual input in small dialogues slightly hindered comprehension.

Wagner (2007) examined the same issue but focused on the influence of text type on the way subjects process visual input and compares the time spent watching the monitor screen of 36 subjects when academic lectures and dialogue videos are played in the listening test. Overall, subjects looked at the monitor screen 69% of the time when the video was being played but watched the dialogue (72%) longer than the academic lecture (67%). Wagner believes that the reason is that dialogue is the interaction between two speakers, with a high degree of contextual dependence, so the contextual cues and non-verbal information are numerous and significant, which are very helpful for learners' understanding. In contrast, lectures are less context-dependent, less interactive, and non-verbal information is poor and unclear. Suvorov's results are consistent with this. He found that the use of video in dialogue had little effect on students' listening comprehension, but the use of video in lectures hindered comprehension (Suvorov, 2009).

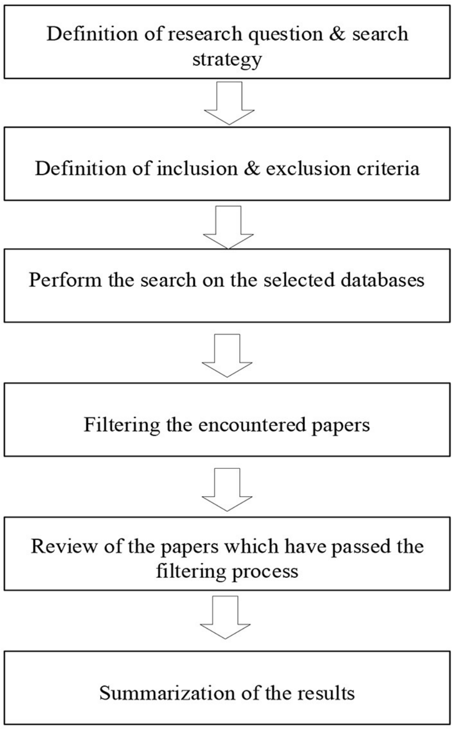


FIGURE 1
Steps of SLR in this study.

Personal factors

The third factor influencing the difficulty of audiovisual multimodal input is personal factors. Currently, studies that specifically examine personal factors in visual input processing are rare. However, the results of many studies have incidentally found that the personal factors of the subjects, such as learning style, cultural background, language level, etc., may affect their performance in the listening test with visual input (Ockey, 2007; Rajabi et al., 2021).

In a study by Fay and Buchweitz (2014), the hypothesis of the study was that personal factors in working memory capacity of L2 learners would predict listening comprehension performance in a proficiency test. The experiment was conducted in two stages, and the participants included 24 students. In the first part of the experiment, 24 students were given a listening test. In the second part of the experiment, 24 students were tested for working memory span. The experimental results show that larger the working memory storage capacity is, the higher the scores in listening comprehension tasks will be.

A study by Masrai (2020) was conducted among 130 non-native English speakers and examined how much of the differences in listening comprehension were explained by auditory vocabulary knowledge, written vocabulary knowledge, and working memory capacity. Results showed that auditory vocabulary knowledge was the strongest predictor of listening comprehension, followed by working memory ability, while written vocabulary knowledge contributed little. This study discusses the influence of auditory vocabulary knowledge and working memory on the interpretive power of listening comprehension and teaching practice in the second language classroom.

To sum up, existing research has proved that visual input has an impact on the performance of second language learners on listening tests, but whether the impact is positive or negative, and the extent of the impact is still inconclusive. Research has also begun to focus on the interaction between visual input and other factors, such as text type and personal factors, but such research identifying influential factors in this interaction is still in its infancy.

Systematic review protocol

This section outlines the research methodology and research process as well as the screening criteria in the literature review. This research mainly focuses on the influence of audio-visual multimodal input on second language acquisition and the influencing factors of the difficulty of second language audio-visual multimodal input. Figure 1 illustrates the systematic literature review process used in this study.

The systematic literature review is launched before March 2022, and the collected papers are also the literature from the 10-year period back to 2012 before that. We plan to begin our review by formulating research questions, defining a search strategy, and keywords for search. When defining our search strategy, we will also define inclusion and exclusion criteria, which will tell us what types of studies we should include in our study. We will then perform our search in the database to obtain relevant studies based on our keywords. The keywords will typically result in a set of papers that may or may not be relevant to our study, so we will need to narrow the list by filtering only those studies that may be useful or relevant to our study. Afterward, we will start reviewing these papers and summarize the results based on the analysis performed. The following topics describe these steps in more detail.

Search strategy

In this section, the approach used for finding the relevant studies to answer the research questions is presented.

Data source

All the retrieved journals are from the following databases: Google Scholar, Eric, Scopus, and Wiley Library. It is not easy to find the literature among the many pieces of literature. Here, the keyword index is mainly used to find relevant documents, the search scope is also expanded through the replacement of synonyms, and the secondary search is carried out through the relevant documents.

Search terms

The keyword search method to search for relevant literature was used. The keywords were generally selected from the title and abstract and are related to the research objectives of the paper. In the advanced search, two to three keywords were selected at the same time to search side by side, because the focus of each database may be different. For this article, the following search terms were used in performing data searches: audiovisual input, or video input, multimodal listening, video-based, visual-based.

Search procedure

Firstly, the data related to this research topic was searched in the paper database, and all papers were screened in three stages. Filter 1 was based on title and abstract keywords of published papers, and studies unrelated to research are excluded. To further refine the results, filter 2 was used, excluding some irrelevant literature. Finally, the rest was filtered for primary studies based on overall quality, using filter 3. For the search procedure, we followed the guidelines provided by Kitchenham et al. (2009) and the entire search process is summarized in Figure 2.

Study selection

Inclusion criteria

The study was selected to find a paper that was relevant to the research question. A key criterion was that the study must focus on the topic of English listening.

The following inclusion criteria were developed:

1. The study had to be a research paper that had been published in peer-reviewed journals and conferences.
2. This was based primarily on the databases Google Scholar, Eric, Scopus, and Wiley online library.
3. The study must be relevant to the research question.
4. The research should be available in open access and full-text format.
5. The research should be published between 2012 and 2022.

Exclusion criteria

The following are the exclusion criteria:

1. Studies not written in English.
2. Papers, reports, books. Studies that are not defined as reliable (e.g. web pages).
3. Studies that are not related to our research.
4. Studies that are not accessible.
5. Those studies that are duplicates.

Quality assessment

Quality assessment (QA) of reviewed literature is paramount to a Systematic Literature Review (SLR) as the quality of conclusions completely depends on the quality of selected literature. Quality Assessment assessing the quality of evidence contained within a systematic review is as important as analyzing the data within. Results from a poorly conducted study can be skewed by biases from the research methodology and should be interpreted with caution. Selecting an appropriate tool to help analyze strength of evidence and imbedded biases within each paper is also essential. If using a systematic review manuscript development tool (e.g., RevMan), a checklist may be built into the software. Other software (e.g., Rayyan) may help with screening search results and discarding irrelevant studies. The following tools may help with study assessment and critical appraisal. The Table 1 below is the specific content of this quality assessment review. The quality assessment team consists of five professors, all of whom are language experts from colleges and universities in China. They will be trained before the review to let them understand the quality assessment standards. There are four options for the standard, [ ] Yes [ ] No [ ] Can't tell [ ] N/A. The selected papers to be reviewed must all meet the review conditions before they can be included in the category of systematic literature review. Only papers with all options of YES can be used. This evaluation team spent two weeks to finally screen out 46 materials from 169 papers that meet the review criteria and can be included in the systematic literature review.

Data extraction

From the keywords search criteria, a total of 12,664 articles were retrieved. After checking the title and abstract, 12,308 papers that did not match the theme were deleted, and leaving only 356 papers. After considering the inclusion and exclusion criteria, the search procedure was further narrowed to 169 papers, and the final papers were evaluated after the quality assessment of the papers. The literature scope was determined to be 46 articles. The PRISMA 2020 flow diagram for systematic literature reviews is shown in Figure 3 below.

In order to extract the required data for further systematic literature analysis studies, the detailed research content and research gaps of each study were assessed. Literature selected in this paper were collected and searched on related topics from

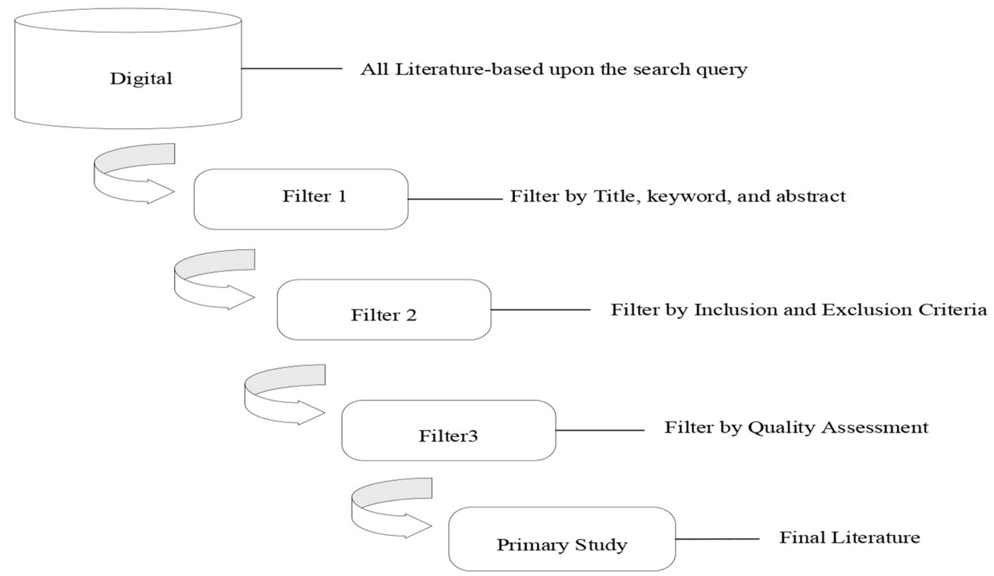


FIGURE 2
Search process.

the decade 2012 to 2022. They were preliminarily classified by topic then recorded in the excel sheet, and the items include search engine, item type, publication year, author, title, abstract, research significance, research object, research question, research limitation, research results, etc. Overall Search Result According to the Search Engine are shown in Table 2 below.

Findings and discussion

What are the effects of audiovisual multimodal input on second language listening comprehension?

With the continuous emphasis on multimedia teaching reform, the research on embedding video teaching in listening courses has been increasing in the past few years, and the benefits of video teaching have become increasingly unified. Studies have shown that there is a "compensatory mechanism" in listening strategies (Field, 2004), that is when learners are hindered in listening comprehension due to insufficient language knowledge, they often resort to some compensatory information, such as pictures, videos, and text annotations that can be referred to as well as relevant clues such as cultural information, encyclopedic knowledge, and common sense of life that can be extracted from the listener's own mind. Video, as a compensation mechanism in listening strategy, confirms the possibility and rationality of audio-visual integration from different angle. In fact, compared with pure listening, audio-visual can promote second language learners to use the top-down listening comprehension mode more to make up for the lack of language knowledge (Mohsen, 2016; Pardo-Ballester, 2016). Moreover, compared with pure audio, audio-visual texts can provide more authentic and vivid language input and more adequate and richer multimodal cultural and situational contexts, which can better promote learners' understanding of the content and have a better understanding of the content (Batty, 2015; Lesnov, 2017; Hsieh, 2020).

In terms of the difficulty of listening, some of the research show that audio-visual materials are less difficult than pure audio materials of the same content, and video has a greater role in promoting understanding than audio; audio-visual input can improve the second language learners' understanding of the material text and can promote the development of listening skills. For example, Jaqueline (2019) found that students who trained listening through video stories made faster progress than those who didn't have visual aids. Ockey and Wagner (2018) found that in foreign language learning, the listening comprehension ability of classes with videotaped instructional materials under the guidance of teachers was significantly higher than that of traditional teaching methods. However, in the process of listening comprehension, the use of multimodal input methods needs to consider the connection between cognitive limitation and working memory (Batty, 2021). It is also true that learners are able to process information from

TABLE 1 Criteria used in quality assessment of systematic reviews.

1. Is a focused multimodal listening clearly stated?	[ ] Yes [ ] No [ ] Can't tell [ ] N/A
2. Are the search methods used to identify relevant studies clearly described?	[ ] Yes [ ] No [ ] Can't tell [ ] N/A
3. Was a comprehensive literature search performed?	[ ] Yes [ ] No [ ] Can't tell [ ] N/A
4. Was selection bias avoided?	[ ]Yes [ ] No [ ] Can't tell [ ] N/A
5. Was there duplicate study selection and data extraction?	[ ]Yes [ ] No [ ] Can't tell [ ] N/A
6. Were the characteristics of the included studies provided?	[ ]Yes [ ] No [ ] Can't tell [ ] N/A
7. Was the scientific quality of the included studies assessed and documented?	[ ] Yes [ ] No [ ] Can't tell [ ] N/A
8. Were the methods used to combine the findings of studies appropriate?	[ ] Yes [ ] No [ ] Can't tell [ ] N/A
9. Was the scientific quality of the included studies used appropriately in formulating conclusions?	[ ] Yes [ ] No [ ] Can't tell [ ] N/A
10. Was publication bias assessed?	[ ] Yes [ ] No [ ] Can't tell [ ] N/A
11. Was the conflict of interest stated?	[ ] Yes [ ] No [ ] Can't tell [ ] N/A
12. Are the stated conclusions supported by the data presented?	[ ] Yes [ ] No [ ] Can't tell [ ] N/A

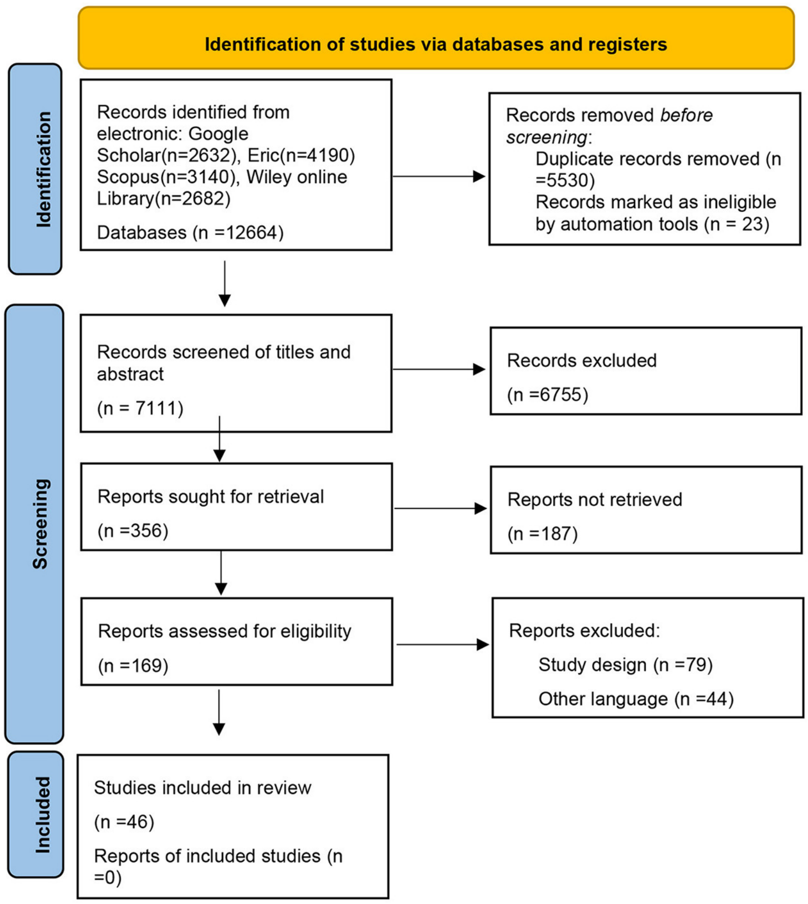


FIGURE 3
PRISMA 2020 flow diagram for systematic literature reviews.

multiple sources simultaneously, and when multimodal inputs are properly integrated, learning is most often beneficial (Rogowsky et al., 2016; Bozorgian and Alamdari, 2018). Hence, adding visual modality information to listening comprehension tasks can not only train listeners to increase the capacity of working memory by simultaneously activating audio-visual channels but also help focus their attention to what is important.

TABLE 2 Overall search result according to the search engine.

No	Search engine	No. of Initial result	Filter1	Filter2	Filter3
1	Google Scholar	2,632	89	45	20
2	Eric	4,190	59	22	9
3	Scopus	3,140	129	56	15
4	Wiley online Library	2,682	79	46	2
	Total	12,664	356	169	46

What are the key factors that affect listening test scores when using visual input?

The use of subtitles

In visual input, Bairstow and Lavaur (2012) pointed out that subtitles are an important feature of video and important content that affects comprehension. However, They do not discuss the influence of subtitles and subtitle types on audiovisual difficulty but focus on the influence of subtitles on audiovisual understanding. As far as discourse comprehension is concerned, the effect of subtitled audiovisuals is better than that of pure audiovisual. Orero et al. (2018) investigated and compared the effects of three kinds of subtitles on learners' understanding of video content, and found that the subjects who watched the native language subtitles had a better comprehension of the video content than those who watched the target language subtitles (Karakas and Sariçoban, 2012; Winke et al., 2013; Birulés-Muntané and Soto-Faraco, 2016; Ebrahimi and Bazaee, 2016). The test group with target language subtitles was better than the test group without subtitle assistance (Karakas and Sariçoban, 2012; Alabsi, 2020).

After investigating the influence of multimodal input on the listening comprehension of English majors, Lin (2016) found that the multimodal combination of video with English subtitles promoted the students' listening content comprehension the most, followed by pure audio, and the effect of the modality of video with Chinese subtitles are minimal. Lesnov (2022) found that keyword subtitles promote audio-visual comprehension and enhance vocabulary acquisition more than full subtitles. Scholars basically agree that the existence of subtitles can promote audio-visual understanding, but we cannot infer that the existence of subtitles reduces the difficulty of audio-visual (Zhyrun, 2016; Bougiatiotis and Giannakopoulos, 2018). Perhaps as pointed out by (Napikul et al., 2018; p. 158), "reading subtitles may interfere with listening comprehension." The learner's audio-visual comprehension is likely to be enhanced by reading subtitles, not by audio-visuals alone. Nevertheless, the use of subtitles can be considered a key factor in listening comprehension but the actual influence of subtitles on audio-visual difficulty needs to be further studied (Leveridge and Yang, 2014; Hsieh, 2020).

Nature of the visual input

Different types of video images have different effects on learners' second language learning (Gilakjani, 2012; Al Mamun, 2014; Zhang et al., 2017; Winarto et al., 2020). Research so far has largely focused on the impact of two types of video images on visual input: one is content visuals, that is, videos that contain salient information images; the other is context-only visuals that only display the speakers' image exemplified in videos such as talk shows and newscasters reading the news. The content video provides a large amount of information input such as pictures, objects, and real scenes (Dehghani and Jowkar, 2012; Pardo-Ballester, 2016; Green, 2017). Compared with context video, it is seen by some to significantly improve the overall understanding level of learners; but it does not help learners understand the uncommon words they listen to and may even interfere with phonological and vocabulary memory (Gathercole and Baddeley, 2014; Wen et al., 2015).

Gabeur et al. (2020) also argues that the content video input provides more information that is beneficial to listeners' understanding to a certain extent because the scene video only presents the image of the speaker or a fixed scene, the learning process is relatively rigid. Some studies have shown that the close-up of the speaker's head (talking head) can hardly provide additional information to promote the listener's understanding (Crook and Schofield, 2017; Hamdan and Al-Hawamdeh, 2018; Zheng and Samuel, 2019). On the contrary, it may also be possible that the listener at this time tends to focus on pure listening interpretation. Fussalam et al. (2019), for example, found that there was no significant difference in the understanding of the content between the audio-visual and pure listening of the talk show.

Alwehaibi (2015) studied the video recording of the lecture and the comparative input effect of the lecture recording and found that the video group that could see the speaker's facial expressions and body movements had significantly higher scores than the lecture recording group. The same findings have been similarly expressed in studies by other researchers: the learners' comprehension of the lecture content was significantly higher than that of the recording group (Missildine et al., 2013). Both the lecture video and the talk show only presented the image of the speaker, but compared with the pure audio, the experimental results are not consistent (Friesen, 2014; Che et al., 2017). It may be related to the text content of the video in these scenes and the purpose of the video playback. Although the effect of context video in promoting understanding is weaker than that of content video, on the whole, scene video can promote the understanding of language input more than pure audio. Some research shows that video type (news vs. speech) has a significant effect on difficulty perception and audiovisual comprehension (Perez et al., 2014).

Relationship between the audio and visual input

Different types of audio and visual relationships in videos will affect audio-visual comprehension, which in turn affects the judgment of the difficulty of audio-visual material (Mathisen, 2012). A direct audio-visual relationship (meaning that the image and voice explanation have a high degree of semantic relevance) and an indirect audio-visual relationship (partial semantic redundancy between the image and the voice explanation) can promote learners' understanding of listening content; however, the close-up of the speaker's head and the discrete sound and picture relationship not only cannot improve comprehension but can even hinder comprehension.

Wong et al. (2012) pointed out that no matter how the audio-visual relationship is coordinated, the information of the visual modality will more or less interfere with the learners. Therefore, to some extent, the relationship between sound and picture may be a cognitive load for learners (Kalyuga and Sweller, 2014). As pointed out earlier, when learners are faced with a more rigid scene video picture, in order to adjust the cognitive load, they are more inclined to only begin interpreting the input through pure "audio-only" processing to obtain information.

To sum up, we found that although there is abundant research on the influencing factors of audio-visual difficulty, there are still some influencing factors to be further proved, especially the further discussion of audio-visual characteristic factors (Akhtar, and Falk, 2017).

Conclusion

The research on second language audio-visual multimodal input is mainly based on the input hypothesis and cognitive load theory and has a deep understanding of the influence of audio-visual multimodal input on second language acquisition and the factors affecting the difficulty of second language audio-visual multimodal input. In general, compared with traditional single-modal input, audio-visual multimodal input has significant advantages in second language listening comprehension and second language vocabulary acquisition, but what are the advantages of audio-visual multimodal input and how the so-called "multi-modality" should be configured and other issues need further study. SLR research shows that there are some factors affecting audio-visual multimodal input difficulty which are focus on subtitles, different video input, and the relationship between sound and picture, but the influence of language and auditory factors on listening difficulty needs further research. We believe that although the research on L2 audio-visual multimodal input has made great progress, there is still a lot of room for expansion. Grading urgently needs a relatively scientific standard. However, we also believe that with the emphasis on audio-visual multi-modal input and more in-depth exploration in the academic and industrial circles, audio-visual multi-modal input will become the main learning method in second language learning in the future. Moreover, due to the gradual increase in the importance of multimodal teaching methods in the field of second language teaching research, the research methods are also more scientific, and the use of empirical research and technology is becoming more and more extensive.

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

TS contributed to the conceptualization, investigation, and writing—original draft. AS and LI contributed to the conceptualization, writing—review and editing, and supervision. All authors contributed to the article and approved the submitted version.

Acknowledgments

The author is very grateful to various parties who have helped and were involved in data collection and the implementation of this research.

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

Akhtar, Z., and Falk, T. H. (2017). Audio-visual multimedia quality assessment: a comprehensive survey. *IEEE Access* 5, 21090–21117. doi: 10.1109/ACCESS.2017.2750918

Al Mamun, M. (2014). *Effectiveness of audio-visual aids in language teaching in tertiary level* (Doctoral dissertation). BRAC University, Dhaka, Bangladesh.

Alabsi, T. (2020). Effects of adding subtitles to video *via* apps on developing EFL students' listening comprehension. *Theory Pract. Second Lang.* 10, 1191–1199. doi: 10.17507/tpls.1010.02

Alwehaibi, H. O. (2015). The impact of using Youtube in EFL classroom on enhancing EFL students' content learning. *J. Coll. Teach. Learn.* 12, 121–126. doi: 10.19030/tlc.v12i2.9182

Arbab, I. A. E. (2020). *Effect of using audio-visual materials on students' language achievements (A case study of secondary schools students at Eastern Gazeera)* (Doctoral dissertation). Sudan University of Science & Technology, Khartoum, Sudan.

Bairstow, D., and Lavaur, J. M. (2012). "Audiovisual information processing by monolinguals and bilinguals: effects of intralingual and interlingual subtitles," in: *Audiovisual Translation and Media Accessibility at the Crossroads* (Leiden: Brill), 273–293. doi: 10.1163/9789401207812_016

Batty, A. O. (2015). A comparison of video-and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Lang. Test.* 32, 3–20. doi: 10.1177/0265532214531254

Batty, A. O. (2021). An eye-tracking study of attention to visual cues in L2 listening tests. *Lang. Test.* 38, 511–535. doi: 10.1177/0265532220951504

Birulés-Muntané, J., and Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PLoS ONE* 11, e0158409. doi: 10.1371/journal.pone.0158409

Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., Ross, S., et al. (2010). *What Makes Listening Difficult? Factors Affecting Second Language Listening Comprehension*. College Park, MD: Maryland Univ College Park. doi: 10.21236/ADA550176

Bougiatiotis, K., and Giannakopoulos, T. (2018). Enhanced movie content similarity based on textual, auditory and visual information. *Expert Syst. Appl.* 96, 86–102. doi: 10.1016/j.eswa.2017.11.050

Bozorgian, H., and Alamdari, E. F. (2018). Multimedia listening comprehension: metacognitive instruction or metacognitive instruction through dialogic interaction. *ReCALL* 30, 131–152. doi: 10.1017/S0958344016000240

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 363, 1001–1010. doi: 10.1098/rstb.2007.2155

Canning-Wilson, C., and Wallace, J. (2000). Practical aspects of using video in the foreign language classroom. *ITESLJ.* 6, 36–31.

Chao, L., Tao, J., Yang, M., Li, Y., and Wen, Z. (2015). "Long short term memory recurrent neural network based multimodal dimensional emotion recognition," in *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge - AVEC '15 (ACM)*. doi: 10.1145/2808196.2811634

Che, X., Yang, H., and Meinel, C. (2017). Automatic online lecture highlighting based on multimedia analysis. *IEEE Trans. Learn. Technol.* 11, 27–40. doi: 10.1109/TLT.2017.2716372

Cheng, G., and Chau, J. (2016). Exploring the relationships between learning styles, online participation, learning achievement and course satisfaction: an empirical study of a blended learning course. *Br. J. Educ. Technol.* 47, 257–278. doi: 10.1111/bjet.12243

Cohen, A. D. (2014). *Strategies in Learning and Using a Second Language*. **Oxfordshire**: Routledge. doi: 10.4324/9781315833200

Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: a case study. *System* 29, 1–14. doi: 10.1016/S0346-251X(00)0 0057-9

Crook, C., and Schofield, L. (2017). The video lecture. *Internet High. Educ.* 34, 56–64. doi: 10.1016/j.iheduc.2017.05.003

Cubilo, J., and Winke, P. (2013). Redefining the L2 listening construct within an integrated writing task: considering the impacts of visual-cue interpretation and note-taking. *Lang. Assess. Q.* 10, 371–397. doi: 10.1080/15434303.2013.824972

Dehghani, A. P., and Jowkar, M. (2012). The impact of computer-video projector on EFL learners' listening comprehension. *Acad. Res. Int.* 3, 106. Available online at: http://www.jallr.ir/index.php/JALLR/article/view/394/pdf394

Ebrahimi, Y., and Bazaee, P. (2016). The effect of watching English movies with standard subtitles on EFL learners' content and vocabulary comprehension. *J. Appl. Linguis. Lang. Res.* 3, 284–295.

Fay, A., and Buchweitz, A. (2014). Listening comprehension and individual differences in working memory capacity in beginning L2 learners. *Letrônica* 7, 113–129. doi: 10.15448/1984-4301.2014.1.16839

Field, J. (2004). An insight into listeners' problems: too much bottom-up or too much top-down?. *System* 32, 363–377. doi: 10.1016/j.system.2004.05.002

Friesen, N. (2014). A brief history of the lecture: a multi-media analysis. *MedienPädagogik* 24, 136–153. doi: 10.21240/mpaed/24/2014.09.30.X

Fussalam, Y. E., Lestari, R., and Anggelia, R. Y. (2019). A study of listening skills through movie: a review of the current literature. *J. Lang. Educ. Dev.* 1, 158–168.

Gabeur, V., Sun, C., Alahari, K., and Schmid, C. (2020). "Multi-modal transformer for video retrieval," in *ECCV 2020 - European Conference on Computer Vision, Aug 2020* (Glasgow), 214–229. doi: 10.1007/978-3-030-58548-8_13

Gathercole, S. E., and Baddeley, A. D. (2014). *Working Memory and Language*. London: Psychology Press. doi: 10.4324/9781315804682

Gilakjani, A. P. (2012). The significant role of multimedia in motivating EFL learners' interest in english language learning. *Int. J. Mod. Educ. Comput. Sci.* 4, 57–66. doi: 10.5815/ijmecs.2012.04.08

Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Lang. Test.* 19, 133–167. doi: 10.1191/0265532202lt225oa

Green, R. (2017). "What is involved in assessing listening?" in *Designing Listening Tests* (London: Palgrave Macmillan), 1–26. doi: 10.1057/978-1-349-68771-8_1

Gregg, K. R., and Krashen, S. D. (1986). The input hypothesis: Issues and implications. *TESOL Q.* 20, 116. doi: 10.2307/3586393

Hamdan, J. M., and Al-Hawamdeh, R. F. (2018). The effects of 'Face'on listening comprehension: evidence from advanced Jordanian speakers of English. *J. Psycholinguist. Res.* 47, 1121–1131. doi: 10.1007/s10936-018-9582-7

He, L., Jiang, D., Yang, L., Pei, E., Wu, P., and Sahli, H. (2015). "Multimodal affective dimension prediction using deep bidirectional long short-term memory recurrent neural networks," in *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge* (ACM), 73–80. doi: 10.1145/2808196.2811641

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Comput. Assist. Lang. Learn.* 33, 567–589. doi: 10.1080/09588221.2019.1577898

Hu, Y. J., and Zhang, D. L. (2013). Experimental Study on multimodal Function in Listening Teaching for English majors. *Foreign Lang. World* 5, 20–25. Available online at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2013&filename=WYJY201305005&uniplatform=NZKPT&v=5tYW53MKDlstf0DT3HDvqmvSoqpMQuK5amkIqqUupWD5dR94iWTX3klcB-LU0xtR

Jaqueline, S. H. I. (2019). Effect of input mode on EFL free-recall listening performance: a mixed-method study. *Eurasian J. Appl. Linguist.* 5, 1–22. doi: 10.32601/ejal.543773

Kalyuga, S., and Sweller, J. (2014). *The Redundancy Principle in Multimedia Learning*. Cambridge: The Cambridge Handbook of Multimedia Learning. p. 247. doi: 10.1017/cbo9781139547369.013

Karakas, A., and Sariçoban, A. (2012). The impact of watching subtitled animated cartoons on incidental vocabulary learning of ELT students. *Teach. Engl. Technol.* 12, 3–15. Available online at: https://files.eric.ed.gov/fulltext/EJ1144953.pdf

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., Linkman, S., et al. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Inf. Softw. Technol.* 51, 7–15. doi: 10.1016/j.infsof.2008.09.009

Lesnov, R. O. (2017). Using videos in ESL listening achievement tests: effects on difficulty. *Eurasian J. Appl. Linguist.* 3, 67–91. doi: 10.32601/ejal.461034

Lesnov, R. O. (2022). Furthering the argument for visually inclusive L2 academic listening tests: the role of content-rich videos. *Stud. Educ. Eval.* 72, 101087. doi: 10.1016/j.stueduc.2021.101087

Leveridge, A. N., and Yang, J. C. (2014). Learner perceptions of reliance on captions in EFL multimedia listening comprehension. *Comput. Assist. Lang. Learn.* 27, 545–559. doi: 10.1080/09588221.2013.776968

Lin, L. F. (2016). The impact of video-based materials on chinese-speaking learners' english text comprehension. *English Lang. Teach.* 9, 1–13. doi: 10.5539/elt.v9n10p1

Masrai, A. (2020). Exploring the impact of individual differences in aural vocabulary knowledge, written vocabulary knowledge and working memory capacity on explaining L2 learners' listening comprehension. *Appl. Linguist. Rev.* 11, 423–447. doi: 10.1515/applirev-2018-0106

Mathisen, P. (2012). Video feedback in higher education–a contribution to improving the quality of written feedback. *Nord. J. Digit. Lit.* 7, 97–113. doi: 10.18261/ISSN1891-943X-2012-02-02

Missildine, K., Fountain, R., Summers, L., and Gosselin, K. (2013). Flipping the classroom to improve student performance and satisfaction. *J. Nurs. Educ.* 52, 597–599. doi: 10.3928/01484834-20130919-03

Mohsen, M. A. (2016). The use of help options in multimedia listening environments to aid language learning: a review. *Br. J. Educ. Technol.* 47, 1232–1242. doi: 10.1111/bjet.12305

Namaziandost, E., and Nasri, M. (2019). The impact of using audio-visual aids on teaching listening among Iranian pre-intermediate EFL learners. *LLT J.* 22, 246–259. doi: 10.24071/llt.v22i2.2028

Napikul, S., Cedar, P., and Roongrattanakool, D. (2018). The effects of film subtitles on English listening comprehension and vocabulary. *Int. J. Appl. Linguist. English Lit.* 7, 104–111. doi: 10.7575/aiac.ijalel.v.7n.6p.104

Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Lang. Test.* 24, 517–537. doi: 10.1177/0265532207080771

Ockey, G. J., and Wagner, E. (2018). *Assessing L2 listening: Moving towards authenticity* (Vol. 50). John Benjamins Publishing Company. doi: 10.1075/lllt.50

Orero, P., Doherty, S., Kruger, J. L., Matamala, A., Pedersen, J., Perego, E., et al. (2018). Conducting experimental research in audiovisual translation (AVT): a position paper. *JosTrans*. 105–126. Available online at: http://www.jostrans.org/issue30/art_orero_et_al.php

Pardo-Ballester, C. (2016). Using video in web-based listening tests. *J. New Approaches Educ. Res.* 5, 91–98. doi: 10.7821/naer.2016.7.170

Perez, M. M., Peters, E., and Desmet, P. (2014). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL* 26, 21–43. doi: 10.1017/S0958344013000256

Peters, E., and Muñoz, C. (2020). Introduction to special issue Language learning from multimodal input. *Stud. Second Lang. Acquis.* 42, 489–497. doi: 10.1017/S0272263120000212

Rajabi, P., Mahmoodi, K., and Hosseini, S. A. (2021). Flipped classroom model and its impact on Iranian EFL learners' classroom anxiety and listening performance. *Comput. Assist. Lang. Learn. J.* 22, 1–16. Available online at: http://www.callej.org/journal/22-3/Rajabi-Mahmoodi-Hosseini2021.pdf

Rogowsky, B. A., Calhoun, B. M., and Tallal, P. (2016). Does modality matter? The effects of reading, listening, and dual modality on comprehension. *Sage Open* 6, 2158244016669550. doi: 10.1177/2158244016669550

Seeber, B. U., Kerber, S., and Hafter, E. R. (2010). A system to simulate and reproduce audio–visual environments for spatial hearing research. *Hear. Res.* 260, 1–10. doi: 10.1016/j.heares.2009.11.004

Surguladze, S. A., Calvert, G. A., Brammer, M. J., Campbell, R., Bullmore, E. T., Giampietro, V., et al. (2001). Audio–visual speech perception in schizophrenia: an fMRI study. *Psychiatry Res.* 106, 1–14. doi: 10.1016/S0925-4927(00)00081-0

Suvorov, R. (2009). "Context visuals in L2 listening tests: The effects of photographs and video vs. audio-only format," in *Developing and Evaluating Language Learning Materials,* eds C. A. Chapelle, H. G. Jun, and I. Katz (Ames, IA: Iowa State University), 53–68.

Sweller, J. (2010). "Cognitive load theory: Recent theoretical advances," in *Cognitive Load Theory*, eds J. L. Plass, R. Moreno, and R. Brünken (Cambridge University Press), 29–47. doi: 10.1017/CBO9780511844744.004

Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Lang. Learn. Technol.* 11, 67–86. doi: 10.1016/j.system.2010.01.003

Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Lang. Test.* 27, 493–513. doi: 10.1177/0265532209355668

Wang, C., Yan, J., and Liu, B. (2014). An empirical study on washback effects of the internet-based college english test band 4 in China. *English Lang. Teach.* 7, 26–53. doi: 10.5539/elt.v7n6p26

Wen, Z., Mota, M. B., and McNeill, A. (2015). Working Memory in Second Language Acquisition and Processing. *Multilingual Matters*. doi: 10.21832/9781783093595

Winarto, W., Syahid, A., and Saguni, F. (2020). Effectiveness the use of audio visual media in teaching islamic religious education. *Int. J. Contemp. Islam. Educ.* 2, 81–107. doi: 10.24239/ijcied.Vol2.Iss1.14

Winke, P., Gass, S., and Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: an eye-tracking study. *Mod. Lang. J.* 97, 254–275. doi: 10.1111/j.1540-4781.2013.01432.x

Wong, A., Leahy, W., Marcus, N., and Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learn. Instr.* 22, 449–457. doi: 10.1016/j.learninstruc.2012.05.004

Zhang, S., Zhang, S., Huang, T., Gao, W., and Tian, Q. (2017). Learning affective features with a hybrid deep model for audio–visual emotion recognition. *IEEE Trans. Circuits Syst. Video. Technol.* 28, 3030–3043. doi: 10.1109/TCSVT.2017.2719043

Zheng, Y., and Samuel, A. G. (2019). How much do visual cues help listeners in perceiving accented speech? *Appl. Psycholinguist.* 40, 93–109. doi: 10.1017/S0142716418000462

Zhyrun, I. (2016). Culture through comparison: creating audio-visual listening materials for a CLIL course. *Latin Am. J. Content Lang. Integr. Learn.* 9, 345–373. doi: 10.5294/laclil.2016.9.2.5

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fpsyg.2022.957896

OPEN ACCESS

EDITED BY
Kai-Yu Tang,
National Chung Hsing University,
Taiwan

REVIEWED BY
Te-Lien Chou,
Fu Jen Catholic University, Taiwan
Yun-Fang Tu,
Fu Jen Catholic University, Taiwan

*CORRESPONDENCE
Nagaletchimee Annamalai
naga@usm.my

†These authors have contributed equally to this work

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

RECEIVED 31 May 2022
ACCEPTED 22 August 2022
PUBLISHED 15 September 2022

CITATION
Annamalai N, Ramírez García A, Mažeikienė V, Alqaryouti MH, Rashid RA and Uthayakumaran A (2022) A phenomenological study of online assessment during a pandemic crisis: Insights from Malaysia, Lithuania, and Spain. *Front. Psychol.* 13:957896. doi: 10.3389/fpsyg.2022.957896

A phenomenological study of online assessment during a pandemic crisis: Insights from Malaysia, Lithuania, and Spain

Nagaletchimee Annamalai[1*†], Antonia Ramírez García[2†], Viktorija Mažeikienė[3†], Marwan Harb Alqaryouti[4†], Radzuwan Ab Rashid[5†] and Arulselvi Uthayakumaran[6†]

[1]School of Distance Education (English Section), Universiti Sains Malaysia, George Town, Malaysia, [2]Faculty of Educational Sciences, University of Córdoba, Córdoba, Spain, [3]Faculty of Human and Social Studies, Mykolas Romeris University, Vilnius, Lithuania, [4]Department of English Language, Literature and Translation, Zarqa University, Zarqa, Jordan, [5]Faculty of Languages and Communication, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia, [6]Centre of Modern Languages, University of Malaysia Pahang, Pahang, Malaysia

Many countries, namely, Malaysia, Lithuania, and Spain, shifted to online assessment during the COVID-19 pandemic. This qualitative case study, which involved 18 undergraduate students from the three countries, was conducted to probe insights into their online assessment experience. Data were interpreted from the perspective of the expectancy-value theory of motivation, which focuses on intrinsic attainment, utility, and cost values. The findings revealed that students were motivated to complete their assessment since they experienced flexibility besides having effective assessment guidelines. The positive experiences were related to intrinsic and attainment values; however, the students were also demotivated when stressed, indicating the high-cost value. Utility value was found to overlap with cost value in this study because students were dissatisfied with the online assessment and expressed less preference for this approach in future. This contributes to our understanding that educators must consider utility values when preparing online assessments. The pedagogical implications of this study revolve around the importance of a checklist, mock exams, alternative assessment (Plan B), and video demos for an effective assessment.

KEYWORDS
assessment, COVID-19 pandemic, online learning, student experience, phenomenology

Introduction

COVID-19 was classified as a global pandemic by WHO in March 2020. UNESCO urged higher education institutions to consider remote teaching and online learning. Henceforth, a sudden change was required without much planning, where teachers and students had to become familiar with various online tools and technologies. At the initial stage of the pandemic, many countries postponed their examinations (Butler-Henderson and Crawford, 2020). Still, as the pandemic continued, many higher education institutions decided on alternative assessment and online examination. This was not a common practice for most students, and the motivation to complete the assessment warranted an investigation. Implementing new approaches, methods, and technological tools can be challenging and since the disruption in the assessment practice is a global phenomenon, investigating the "what, when, why, and how" of this phenomenon can provide valuable guidelines to educators and policymakers. It is pertinent for students to be motivated to engage in online assessment. Otherwise, they will not attempt to overcome the difficulties and complete the assessments.

Existing quantitative studies related to online assessment during the pandemic provide little insight into the students' motivation and its importance in online assessment (Khan and Jawaid, 2020; Rahim, 2020). Most studies relied on surveys by using different statements to test a particular model. While these studies ensure many responses and hypotheses testing, they do not capture the nuances afforded by qualitative studies. Tam (2022) calls for more qualitative studies to better understand assessment from the students' point of view.

Also, there has been excessive focus on instructional design, technological, and cognitive aspects of learning with little attention given to emotional aspects (Montero and Suhonen, 2014). If studies are not focused on learners' experience, limited knowledge is gained about the role of students' feelings, needs, and preferences in their online assessments. Failing to explore such pertinent issues "can only offer an incomplete view of the learning experience" (Montero and Suhonen, 2014, p. 165).

The objectives of this study are 2-fold: (1) to examine students' motivation related to online assessment and (2) to put forth the pedagogical implications of an effective online assessment. The findings of this study may inform educators and policymakers in higher education institutions to address the concerns of students and fine-tune assessment initiatives to facilitate students' learning and improve learning outcomes. Our research can contribute to the development of the online assessment, for it adds to the current literature by providing empirical evidence in understanding students' motivation in completing online assessments, especially in the rarely explored context of Malaysia, Lithuania, and Spain. The study focuses on the summative assessment that "occurs at the end of the learning process" (Faulconer et al., 2019, p. 1) where marks, grades, or degrees are given to students (Jones, 1996, p. 134) and enables institutions to make a judgment on one's performance.

The research question for this study is:

1. What motivates the students to complete their online assessment during the COVID-19 pandemic?

Literature review

According to Hannaford (2015), motivational beliefs can influence student assessment. The expectancy-value theory of motivation is a practical framework that guides the assessment task that appears to enhance motivation (Simpson, 2013). The theory was developed by Atkinson in 1950 and further established by Wigfield in 192. The theory emphasizes that motivation is crucial for students to complete a task and value their efforts (Wigfield and Eccles, 2000). It is pertinent for students to have an expectation of success and commit to a task with some positive values. If one has high expectations of success but does not appreciate the task assigned, one will not value the task. In answering the question, "Why should I do this task?," the expectancy-value theory draws our attention to four aspects: intrinsic, attainment, utility, and cost values. Intrinsic value refers to the enjoyment of attempting a task. When learners are intrinsically interested, they become more involved in completing the job. Attainment value details the importance of completing the task well, whereas, utility value is the perception of whether the mission is worthwhile or significant in future. Cost value refers to the effort and commitment to accomplish the task. In general, students tend to experience high-cost value when the assessment is time-consuming and stressful. At the same time, students will experience low-cost value when the online assessment reduces workload, probably by sharing workloads with other students and benefiting from each other.

Review of related studies on online assessment during the COVID-19 pandemic

An exponential number of papers on the COVID-19 pandemic and learning have been published in recent years (e.g., Guangul et al., 2020; Hasan and Bao, 2020; Scull et al., 2020; Annamalai, 2021; Annamalai et al., 2021; Anh, 2022; Yan-Li et al., 2022). These studies attest to how extensively and immensely the research on the COVID-19 pandemic has proliferated. Some studies focus on students' learning experience (Hasan and Bao, 2020), educators' readiness (Annamalai, 2021), teacher stress (MacIntyre et al., 2020), innovation in teaching (Scull et al., 2020; Annamalai et al., 2021), and distance learning during the pandemic (Schneider and Council, 2020).

Guangul et al. (2020) surveyed remote assessment. They identified challenges related to the commitment of the students to submit their review, dishonesty, infrastructure, and assessment of learning outcomes that have been taught. The study suggested that preparing different sets of questions and various types of evaluations will help solve some of the challenges. Sharadgah and Sa'di (2020) examined the preparedness of faculty members for online assessment. The study reported that the educators were not convinced that the online assessment had effectively assessed all the intended learning outcomes. There were also concerns about the lack of experience. It was concluded that many educators are not qualified to conduct the online assessment and they urge higher institutions to consider online invigilation software. White (2020) reported that students hold complex perceptions around their attitudes toward academic integrity and rationalizations of misconduct.

Though studies on online assessment and the COVID-19 pandemic are gaining momentum, it is still unclear whether students are motivated to complete their online assessments during the pandemic. Studies on online assessment during the pandemic remain an understudied area in Malaysia, Spain, and Lithuania. To fill this gap, the study examined students' motivation in completing their online assessment during the COVID-19 pandemic in the context of Malaysia, Spain, and Lithuania.

Methodology

This study employed a qualitative research methodology to gain an in-depth understanding of the event in a real-world setting. A phenomenological approach was chosen as it enables the understanding of the phenomenon of assessment through the eyes of students who are experiencing it. In other words, the attempt was to reveal "What's it like for them?" (Van der Mescht, 2004). Epistemologically, we believe that the phenomenon is socially constructed and given meaning by the individuals experiencing it [see Easterby-Smith et al. (2012)]. The current study used inductive analysis, moving from the concrete to the abstract phases and focusing on non-linear processes that transpired in a natural setting (Lichtman, 2014).

Participants

A total of 18 students (i.e., six from each country) were selected based on a purposive sampling approach but they were recruited conveniently. The six participants from each country were considered for informational purposes rather than statistical deliberations, as recommended by Lincoln and Guba (1985). More importantly, the engagement with the six participants from each country was sufficient for the saturation point, as evidenced by repeated themes. They were third-year students; therefore, they already had a considerable amount of learning experiences to compare the pre–COVID-19 situation and the situation during the lockdown. All the participants were briefed on the nature of the study, confidentiality, and anonymity. The participants were full-time students from various degree programs related to social science. Their fields of study were education, management, counseling, and language. The participants had never experienced online assessment. Before the COVID-19 pandemic, their assessments were often conducted face-to-face in traditional classroom settings.

Data collection and analysis

Semi-structured interviews were conducted with the participants during the COVID-19 *via* Webex or MS Teams. The interviews lasted for 40–60 min for each participant. The semi-structured nature of the discussions allowed the interviewer and participants to explore and discuss issues related to online assessment. As argued by Gilbert et al. (2007), interviews may provide richer insights into the phenomenon being investigated and answers to "why" and "how" questions (p. 571).

The interviews were conducted in the student's native languages—Spanish, Lithuanian, and Malay—and later transcribed verbatim. Translations of the transcriptions into English were done by the three researchers (native speakers of each language) working at the respective universities where the study took place. Two main questions guided the interviews:

1. Can you share your experience related to online assessment during the COVID-19 pandemic?

2. Were you motivated to do your online assessment during the COVID-19 pandemic? Why?

Thematic analysis was employed in this study in analyzing the data. The six-step thematic analysis procedure is: (1) becoming familiar with the data and transcribing all data; (2) generating codes; (3) classifying codes into themes; (4) reviewing and refining themes; (5) concisely defining and naming themes; (6) producing a report from the emerging themes which is a descriptive, analytical, and argumentative narrative. Direct quotations from the participants were included to explain critical themes (Braun and Clarke, 2006, p. 87).

In data familiarization stage, the most common words used were identified. Participants' keywords and main ideas were considered in generating the initial codes. For example, "dissatisfaction" with submission (L3), "not good" assessment (S5), and "cheating" (L4). These keywords were then converted to the code of inefficient assessments. The initial codes were examined repeatedly and categorized according to similar characteristics to search for themes. The themes were checked to see if the initial codes fit into the classification category.

The classification categories were defined and given names indicating their distinct characteristics.

Trustworthiness in qualitative data

The current study was guided by Guba and Lincoln's (1994) four qualitative research criteria: confirmability, credibility, dependability, and transferability. Member checking was used to achieve the concept of credibility (Creswell and Poth, 2017). The transcribed interviews were returned to the participants to determine whether the information provided during the interview was the same as the information in the processed data. A description of the setting and participants enabled transferability. When three experienced lecturers coded the themes, investigator triangulation was used. They were able to reach 90% agreement among the coders. As a result, the findings are trustworthy, convincing, and accurately reflect the actual situation. A panel of experts in technology and educational research also validated the interview questions.

Findings

In presenting the qualitative data, participants from Malaysia were identified as M1, M2, and M3, while participants from Lithuania were L1, L2, L3, and from Spain as S1, S2, and S3.

Flexibility during the assessment

Majority of the participants felt that the "*assessment suited the whole situation (Pandemic), and everyone adjusted to the new normal of education and the flexibility (e.g., L6).*" Similarly, S3 opined that *it is "good that education did not stop because of the lockdown"* (S3). S2 found that the "*assessment was easier than the one conducted under the normal conditions*" (S2). Students were happy because they "*do not have to commute to the exam hall for 90 minutes exam and can sit for exams via laptop and anywhere they want*" (M3). In fact *"some students study even more during the pandemic"* (S6). Students felt *"calm to sit for the examination in the comfort of their home, in their study rooms, without disturbance from other people*" (S6). L3 liked "*the series of quizzes because revision can be done from the feedback*" (L3) and "*the home environment made [him] calmer in a way, at least in some of the examinations."* As for L6, he *"thinks professors should make students more comfortable at the beginning of the exams, reassuring students not to worry about potential technical issues."*

According to M2, the flexibility of the time allowed him to *"research on the questions and understand what is required and it would be clearer for [him] and it would give [him] a whole perspective."* The flexibility in the time allowed them *"to go over the material once again"* (L5). Similarly, M1 expressed that *"the online assessment was less stressful during the pandemic as students were at home and could manage the time division between work and study."* Interestingly, M4 elaborated, "if the pandemic occurred 20 years ago, students would be sitting at home *wondering what to do. Luckily it was 2020, and they have technology tools like Zoom, Google Meet, Microsoft Team, and Skype to communicate with their lecturers."* S6 interestingly pointed out that because of the flexibility of time, her lecturer was able to interact with her; *"she was on maternity leave at that time, but she was commenting on the drafts that I sent to her."*

Students were also contented when they *"received feedback from their professors and allowed them to modify and continue working for better grades."* (S3). The teachers' feedback also helped them "*to complete their subsequent assignment according to their explanations*" (L7), and they were *"able to understand the topic much better"* (M10). **Figure 1** shows the number of participants and categories for the theme related to flexibility in assessment.

Effective assessment guidelines

Some students felt that it was easier to have an online assessment since more guidance was received compared to the face-to-face assessment. The sudden transition to the online assessment made the lecturers guide the students thoroughly. The lecturers gave *a "rubric as a guide because there were many assignments. The rubrics are very useful because it helps students to understand the reasons behind assessment and marks"* (S4*).* With the rubrics, students *"know why students are given certain marks and not the other."* (S4). It is practical to have assessment rubrics because *"I was constantly checking the rubrics for doing each part of the work appropriately, for instance, for completing the theoretical part of the assignments, I checked the rubric to see what had already been done and what had yet to be done"* (S1). Rubrics gave them the idea of *what was expected from [them] and how [they] would be assessed* (S1). Students felt that the university had revamped the assessment and they *"benefitted from it because it helped [them] to pass the subjects"* (S2). For example, *reducing the weight of 40 per cent of the final examination mark so that marks from other assignments can take a bigger percentage* (S2).

S6 highlighted that *"there were many assignments and different professors in the subject, and each part had its share in the final mark."* Similarly, L5 said that *"some professors gave quite many tasks for assessment" and "seven separate essays were done in group work."* Suitable assignments were given to students during the precarious time. Students found that the *"assessments were fun and more participation of students was generated during this pandemic"* (M2). Also, there was a *"requirement to make the assessment criteria more flexible; this helped students achieve better grades"* (S5). **Figure 2** shows the distribution of the number of participants who contribute to the theme of "effective assessment guidelines."

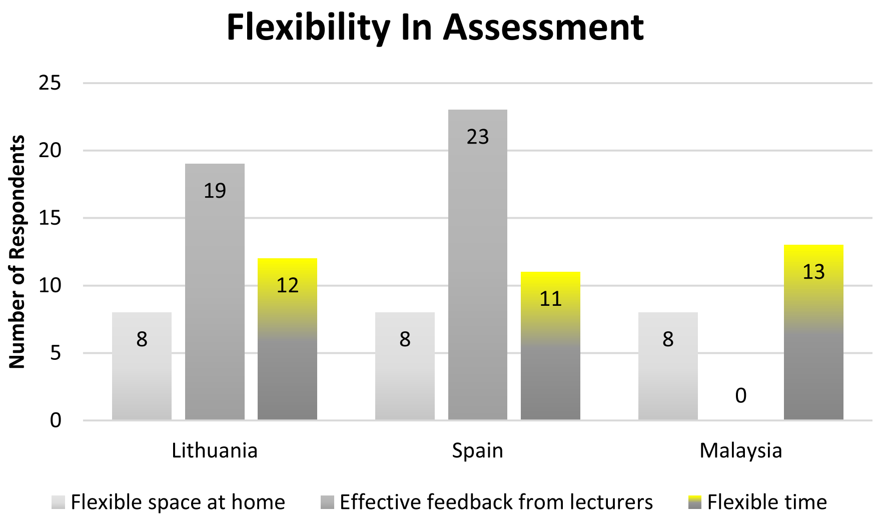


FIGURE 1
Number of the participants and categories contributing to the theme of flexibility in assessment.

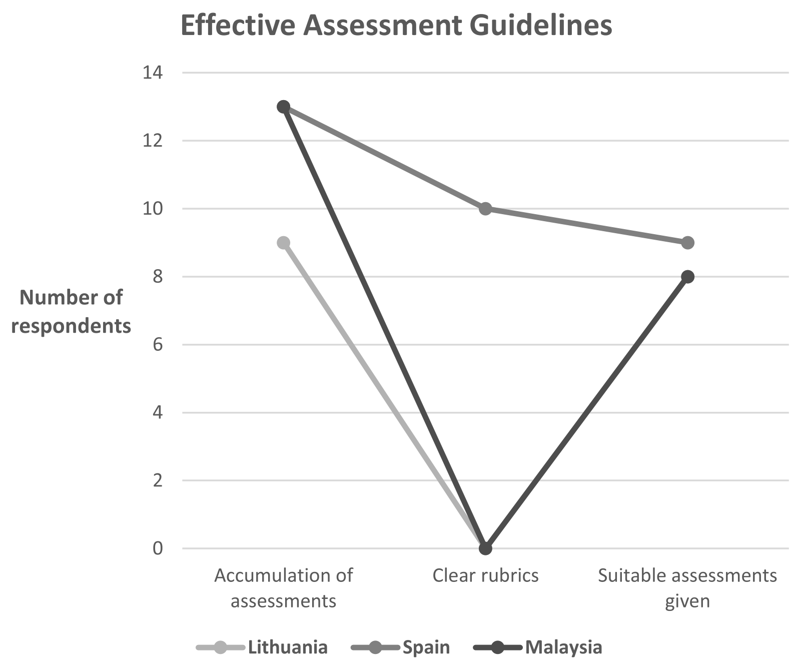


FIGURE 2
Number of participants and categories contributing to the theme of effective assessment guidelines.

Stress and fear of the online assessment

The theme of stress and fear was evident in most of the interviews. The online assessment was stressful for L3 because "*many people were living in the house, and [his] grandparents also came to live with [him], and [they] had to share rooms, the spaces for studying, doing assignments and the exams.*" There was also a situation where students needed *"to agree on who uses the computers depending on [their] classes, assignments, mobile data"* (S5), and *"this has affected the assessment and participation in classes"* (S5). The stress was also caused by *"limited time set during the assessment"* (L4) *and "seeing how much time left"* (L3). Most students were stressed *"when the time was spent solving the technical problem"* (L4). The worst tension was caused by the *"fear of disruption of the Internet connection*

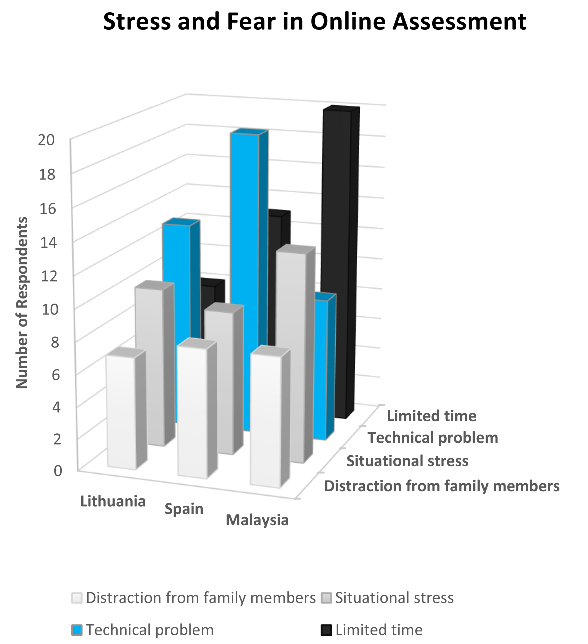


FIGURE 3
Number of participants and categories contributing to the theme related to stress and fear in online assessment.

and by losing time for answering the questions (S2)." L6 expressed her dissatisfaction and detailed that:

> *I did not like the oral part of the examination because I have always experienced much more stress, and now, when we do everything online, the fear is immense. It is mainly because of the Internet connection problem.*

According to M5, "*the laptop [she] was using gets heated up during the assessment when used for a long time and reduces the processor speed. This causes the system to lag.*" One student compared the online assessment and highlighted the advantage of classroom assessment. In her words:

It is less stressful to present in class because even if there is a problem with the slides or something else, you can continue with your presentation. At the same time, you can get cut off entirely during an online production (L2). **Figure 3** shows the number of participants and categories which contribute to the theme related to stress and fear in online assessment.

Inefficient assessment

Lecturers focus on preventing the students from cheating rather than re-thinking their assessment methods and exam tasks (L6, S2, S5). Therefore, lecturers limited the time for the assessment. This led to *"stress levels for those who work slower, like me*" (L5) and before the pandemic, "*I had a very similar subject with the same lecturer. We wrote the examination in class, and for the writing, part was given an hour and a half"* (L5).

When confronted, the lecturer *"explained their decision and expressed the fear for students to cheat while writing from home when they have more time*" (S2). One student *"submitted a complaint to the Dean's office about the unfair time limit. She was unhappy about the teacher's decision to set a limit for answering a question in Moodle."* L4 said, *"I liked writing when we were given more time, for example, one week or two for essay writing, not just one hour as in the exam."* (L4). L1 also commented that:

> *The time was short for 30 tasks. We had only 15 minutes. It was impossible to concentrate on the tasks because students were thinking about how much time was left, not about making tasks correctly.*

M2 gave a similar account and gave the example of *"an Accounting subject that is usually for three hours and shortened to two hours during the online assessment, thus creating panic whereas three hours is the proper designated time needed to answer the questions as the answers need lengthy and thorough working."* Participants highlighted that there was *"no standardisation during the online assessment and felt it was all chaos and there was not a single moment when [they] knew for sure what [they] had to do for some of the subjects (S2)."* There were instances where the lecturer *"added new requirements after the assignment was submitted*" (L2). Sl added that *"coordination from the professors should be planned so that the deadlines would not overlap because this stressed [them] a lot*" and *some initial instructions that the students must follow to complete their assessment successfully* (L2*)*. S2 lamented, "*I was spending all my time on a diversity of assignments, resulting in less time for studying.*" A student said they *"sit in front of the computer too long because some professors gave quite many tasks for assessment, and [she] was exhausted and waiting for the quarantine to end*" (L1). There was a situation when the students *"have back-to-back exams, and there is no time to revise and remember important points."* M2 felt "*it was ridiculous to answer nine essay questions from various subjects and eight separate online MCQ questions in six days.*" Similarly, M5 said that *"the idea of a fixed date and time to answer questions from various places cause a problem, and the university should be aware of problems faced by the students."*

Although students have highlighted the advantages of a flexible environment for assessment, some participants were not satisfied with such evaluations and highlighted the possibilities for students to cheat. For example, *"it could have been done by another person, their father, their mother*" (S6*)*. Furthermore, *"the changes in the percentage for examination and when the exam is 40 per cent, and there is no minimum score to pass, it results in that everyone passes without acquiring the knowledge that has to be acquired*" (S6). One technique that the participant disliked was the instrument in the assessment. In S6's words:

> *Every student had to submit in the same folder on Moodle. The deadline was set. And the teacher pointed out that they would review the works if submitted in advance and tell what could be improved before the deadline. Some students presented their papers in advance and were expecting some reflection. However, the teacher reviewed the works only after the deadline passed and expressed their dissatisfaction with some of them. The students were upset because they could have worked on their papers earlier had the teacher reviewed them.*

L3 laments that she *"was informed by the defence committee to look straight in the camera more, but [she] was more focused on the PowerPoint presentation so that [she] could follow what [she] was going to tell next and to explain things."* She emphasizes that if it was *done "face-to-face, [she] would have indicated at the PowerPoint and would have looked at the committee members more. It is tough to present the Final Thesis orally, to look in the camera and not at the slides."*

As for S5, one of the lecturers *"used a chat in the forum as a type of participation."* S5 noticed that *"it was not a good form of assessment, because people repeat what the first or second students have written and write something like "I agree with XXXX or XXXX" without adding anything new. [She] don't think this is a good assessment tool. But [she] had no doubts that [she] did not like the examination."* (S5). L4 concurred in this respect and explained her dissatisfaction:

> *Sometimes students got lower marks and no explanation was given. A good friend said she tried to prove to the teacher that she did not plagiarise. The teacher said that she checked our work with software and returned it to the student. But the student said she was just reading things online, and maybe that was why the ideas were similar and the sentences were identical.*

As for L1, presentations were complicated because of the *"requirements given by the lecturers. For example, some of the lecturers required students to have their cameras on. If the student has no such opportunity, professors made their grades lower, even if the presentation is well prepared."* L2 felt comfortable presenting in class for assessment because "*it is much easier to control [her] hands and maintain eye contact."*

M1 was disappointed when the percentage for assessment changed due to the pandemic. M1 explained that *"the lecturer decided to change the portion of the project work which was 10% before the pandemic and now it is said to be 20% since the project work is completed. The percentage for the online assessment will be much lower, which will affect [his] final grade. If [he]knew that the contribution was higher, [he] would have put in the extra effort."* M3 commented on the *"one-time click assessment, and after some time when students realised the mistakes and wanted to change, they are not able to do so. Students can still rub the wrong answers if the exam is in the classroom."* **Figure 4** illustrates the number of participants and categories for the theme related to inefficient assessment.

Discussion

Adopting the perspective of the expectancy-value theory of motivation, the interview findings are discussed based on the four main domains of the theory: attainment, intrinsic, cost, and utility values. The participants confirmed that the accumulation of grades for the assessment, clear rubrics, and suitable assessments motivated them to complete their online assessment. These aspects are related to attainment values. With clear rubrics, students could identify expectations and standards for a particular assessment. This, in a way, motivates students to improve their grades. According to Tam (2022), educators' guidance in explaining various types of examination questions help learners in preparation for their assessment. As revealed by the students, their teachers tend to be more emphatic and sensitive in carrying out online assessments. They seem to understand that students may be temporarily unavailable during the pandemic due to health issues, having other responsibilities such as taking care of their family members, having limited access to technology tools etcetera. In such a situation, Doucet et al. (2020, p. 8) proposed the concept of "Maslow before Bloom" indicating that while teaching and learning activities are important, health and safety must take precedence during a crisis. Nguyen et al. (2020) assert that switching to frequent assessment helps students manage their time wisely and break up their study time into smaller segments. Such periodic assessment gives a lower stake and lowers the stress environment. Incorporating various assessments encourages learners to change their approach from traditional assessment to deep and proactive learning (Jacobs, 2021). The findings also support Felder and Brent's (2003) idea that assessment supports learning and leads to constructive educational changes. For example, the project work, online presentation, and task-based education evident in the findings allow student-centered learning and active learning to occur.

The domain of intrinsic value is related to flexibility in assessment. The subcategories related to this theme are flexible time, flexible space, and effective feedback from lecturers. The students appreciate the effective feedback from lecturers, encouraging them to promote and maintain their motivation (Benson and Brack, 2010), since well-developed feedback can assist learners who face learning difficulties (Wang, 2014). Students at University S in this study received clear rubrics, but students at Universities L and M did not. Also, feedback from the lecturers was experienced by students from Universities S and L. However, students in University M were not provided with feedback. Regardless, the students were motivated to complete the assessment since they had flexible time and space

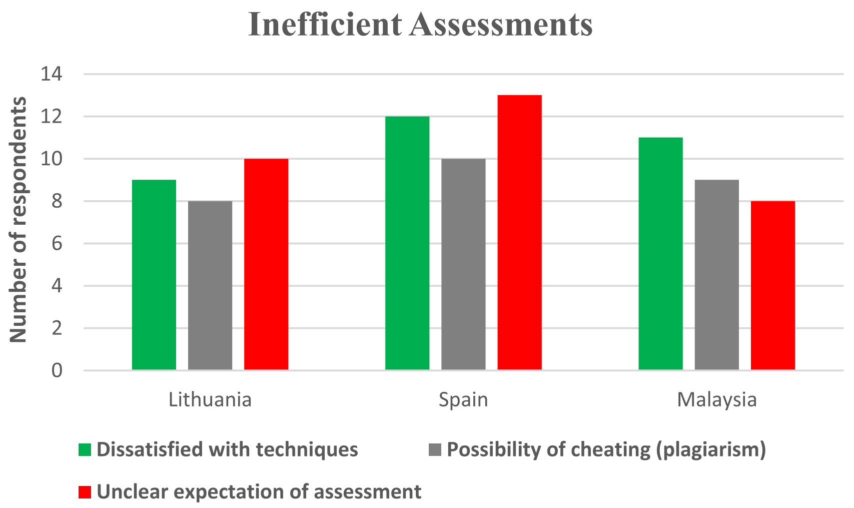


FIGURE 4
Number of participants and categories for the theme related to inefficient assessment.

to complete the assessment. The "space" in which students learn can affect their emotions and engagements (Kayumova and Tippins, 2016). Technology tools resolved the lack of physical attendance, and students were more comfortable and such an environment increased learning outcomes, as affirmed by Goh and Sandars (2020) and Khan and Jawaid (2020). These experiences led students to positive attitudes toward online assessment.

High-cost value was identified when students were demotivated because they experienced distraction from family members, situational stress, technical problems, and limited time to complete the assessment. This is consistent with Tam (2022, p. 12), who argues that "time and technical constraints are inappropriate ways of encouraging students to the fullest with the topics and the assessment process but impose overwhelming pressures on students." It is also evident from the interview that students were not consulted when changes were made in the percentage of the carry marks for the assessment. Some assessments were also scheduled simultaneously for various subjects (asynchronous online assessment) resulting in students experiencing stress and not performing well in their assessments. Hence, the training of educators is an essential component that institutions of higher learning can no longer depreciate but need to be more rigorous, robust, and advanced. As Boud and Dochy (2010) postulate, assessment practices significantly enhance and enrich university students' learning experiences because the improvement in assessment practices impacts the quality of students' learning. Practitioners and educators should have related knowledge, skills, and short-term and long-term planning strategies.

The theme related to utility value is not evident in the findings. Since the COVID-19 pandemic has changed the assessment design (techniques of assessment, level of difficulties, students learning strategies and preparation for assessment), learners should be informed on how such assessment can prepare them for future challenges. For example, the importance of communication, collaboration, critical thinking, and creativity should be explained to the students for them to realize the significance of such assessment. According to Gibbs (2010), it is through assessment design that educators create a healthy learning milieu. At the same time, the students also interpreted the online assessment negatively where they felt that online assessment is not very valuable to them because of their dissatisfaction with the techniques, the possibility of cheating among students, and inconsistent assessment expectations. Therefore, there is an overlap between cost value and utility value. **Figure 5** illustrates the categories and themes related to online assessment examined in this study.

Pedagogical implications

The shift to online assessment due to the COVID-19 pandemic is something unexpected. There are challenges and there is no one-size-fits-all approach for the online assessment. However, the findings from the three countries put forth in this study could guide educators in planning their online assessments. Understanding the problems and considering the appropriate strategies are essential to maintain quality in assessment.

Five recommendations can be made from this study, which are:

1. A checklist should be given to teachers and students to support and mitigate the limitations of an online assessment. To develop the list, educators should consult

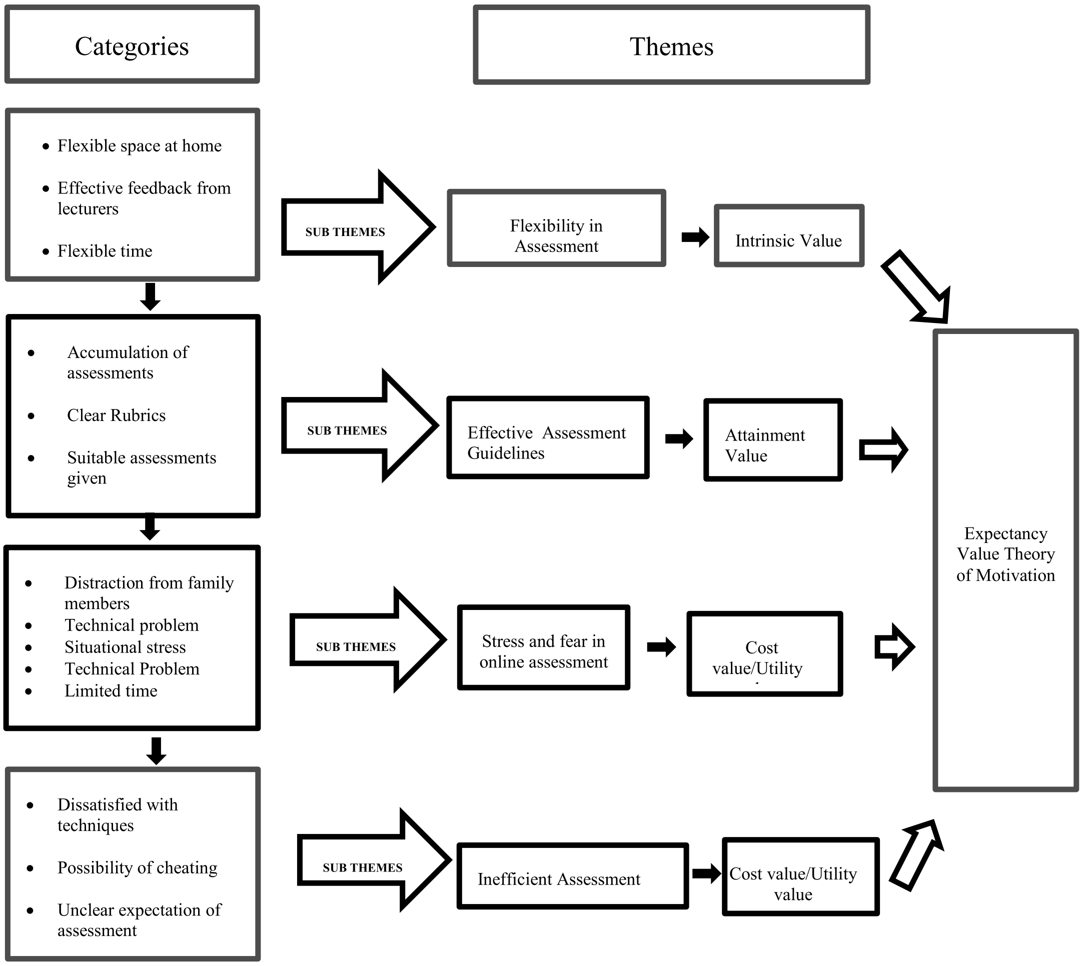


FIGURE 5
Categories and themes related to online assessment.

the students and consider the strengths and regulations they will experience during the online assessment. More importantly, in designing the checklist there should be a discussion and negotiation between educators and students so that consensus can be reached. A checklist detailing how the assessment is beneficial to the students should be included to emphasize utility value. The checklist should also include rubrics and feedback.

2. Mock assessment should be conducted to make students aware of the difficulties and consequently work together to overcome the challenges.
3. A discussion on alternate assessments can be arranged if students face technical difficulties. There must always be "Plan B" when the prepared online assessment fails. Plan B will only be executed when students provide evidence (screenshot of technical problems) to avoid dishonesty.
4. Video demos on how assessment could/should be conducted ought to be created to guide students. The video may increase students' acceptance of online assessments as they would be accustomed to the format and thus have/gain confidence in the system.
5. A well-planned schedule should be created to avoid overlaps in assessment. Some of the assessments are conducted for more than a week (asynchronous online assessment) and it overlaps with other subjects.

Conclusion

The COVID-19 pandemic has opened venues for online assessment with a completely new experience for educators and learners. This study reveals several significant findings related to assessment and motivation. The use of the virtual environment for online assessment comes with advantages and disadvantages. It is a must for students, instructors, and e-learning policymakers to consider the potential benefits and challenges when encouraging e-learning assessment. Future studies should explore and further investigate and experiment with effective online assessment practices that would contribute to students' productive and meaningful learning, especially in times of distress, tragedies, and pandemics, as we are currently experiencing. Also, similar comparative studies involving more countries (developed or developing) should be initiated. The themes derived from this study could also be used to develop constructs (and eventually a questionnaire) that is comprehensive and complete to understand online teaching and learning from both the educators' and the students' perspectives. The third surge of COVID-19 worldwide means that, most likely, the online assessment may last longer than expected. It also means that teaching and learning in higher institutions should be brave to explore, create, develop, implement, and evaluate new assessment modes that would be feasible for educators and students.

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 author.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved 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

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

Anh, V. T. K. (2022). English teachers' attitude and challenges in facing immediate online teaching: A case study in vietnam. *J. Nusantara Stud.* 7, 495–511. doi: 10.24200/jonus.vol7iss2pp495-511

Annamalai, N. (2021). Online learning during COVID-19 pandemic. Are malaysian high school students ready? *Pertanika J. Soc. Sci. Hum.* 9, 1571–1590. doi: 10.47836/pjssh.29.3.06

Annamalai, N., Tangiisuran, B., and Athirah Daud, N. A. (2021). Transitioning to online clerkship during unprecedented times: An innovative online flipped in-patient clerkship. *Innov. Educ. Teach. Int.* 1–11. doi: 10.1080/14703297.2021.1931406

Benson, R., and Brack, C. (2010). *Online Learning and Assessment in Higher Education*. Cawston UK: Woodhead Publishing.

Boud, D., and Dochy, F. (2010). *Assessment 2020. Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council.

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101.

Butler-Henderson, K., and Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Comput. Educ.* 159:104024. doi: 10.1016/j.compedu.2020.104024

Creswell, J. W., and Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, US: Sage publications.

Doucet, A., Netolicky, D., Timmers, K., and Tuscano, F. J. (2020). *Thinking about pedagogy in an unfolding pandemic*. Available online at: https://www.saspa.com.au/wp-content/uploads/2020/04/2020_Research_COVID-19_2.0_ENG-Thinkingabout-Pedagogy-in-an-Unfolding-Pandemic.pdf (accessed January 30, 2022).

Easterby-Smith, M., Thorpe, R., and Jackson, P. R. (2012). *Management research*. Thousand Oaks, US: Sage.

Faulconer, E., Griffith, J. C., and Frank, H. (2019). If at first, you do not succeed: Student behaviour when provided feedforward with multiple trials for online summative assessments. *Teach. High. Educ.* 26, 581–601. doi: 10.1080/13562517.2019.1664454

Felder, R. M., and Brent, R. (2003). Designing and teaching courses to satisfy the ABET engineering criteria. *J. Eng. Educ.* 92, 7–25. doi: 10.1002/j.2168-9830.2003.tb00734.x

Gibbs, G. (2010). *Using assessment to support student learning*. Leeds: Leeds Met Press.

Gilbert, J., Morton, S., and Rowley, J. (2007). e-Learning: The student experience. *Br. J. Educ. Technol.* 38, 560–573. doi: 10.1111/j.1467-8535.2007.00723.x

Goh, P. S., and Sandars, J. (2020). A vision of the use of technology in medical education after the COVID-19 pandemic. *MedEdPublish* 9:49. doi: 10.15694/mep.2020.000049.1

Guangul, F. M., Suhail, A. H., Khalit, M. I., and Khidhir, B. A. (2020). Challenges of remote assessment in higher education in the context of COVID-19: A case study of Middle East College. *Educ. Assess. Eval. Account.* 32, 519–535. doi: 10.1007/s11092-020-09340-w

Guba, E. G., and Lincoln, Y. S. (1994). "Competing paradigms in qualitative research," in *Handbook of qualitative research*, eds N. K. Denzin and Y. S. Lincoln (Thousand Oaks, CA: Sage), 105–117.

Hannaford, L. (2015). Computational Journalism in the UK Newsroom. *J. Educ.* 4, 6–21.

Hasan, N., and Bao, Y. (2020). Impact of "e-Learning crack-up" perception on psychological distress among college students during COVID-19 pandemic: A mediating role of "fear of academic year loss". *Child. Youth Serv. Rev.* 118:105355. doi: 10.1016/j.childyouth.2020.105355

Jacobs, A. D. (2021). Utilizing take-home examinations in upper-level analytical lecture courses in the wake of the COVID-19 pandemic. *J. Chem. Educ.* 98, 689–693. doi: 10.1021/acs.jchemed.0c00768

Jones, S. H. (1996). Crits—an examination. *J. Art Design Educ.* 15, 133–141. doi: 10.1111/j.1476-8070.1996.tb00660.x

Kayumova, S., and Tippins, D. (2016). Toward re-thinking science education in terms of affective practices: Reflections from the field. *Cult. Stud. Sci. Educ.* 11, 567–575. doi: 10.1007/s11422-015-9695-3

Khan, R., and Jawaid, M. (2020). Technology-enhanced assessment (TEA) in COVID 19 pandemic. *Pak. J. Med. Sci.* 36, S108–S110. doi: 10.12669/pjms.36.COVID19-S4.2795

Lichtman, M. (2014). "Drawing meaning from the data," in *Qualitative research for the social sciences*, ed. M. Lichtman (Thousand Oaks, US: SAGE Publications, Inc), 335–340. doi: 10.4135/9781544307756.n12

Lincoln, Y. S., and Guba, E. A. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

MacIntyre, P. D., Gregersen, T., and Mercer, S. (2020). Language teachers' coping strategies during the Covid-19 conversion to online teaching: Correlations with stress, wellbeing and negative emotions. *System* 94:102352. doi: 10.1016/j.system.2020.102352

Montero, S. C., and Suhonen, J. (2014). "Emotion analysis meets learning analytics: Online learner profiling beyond numerical data," in *Proceedings of the 14th Koli calling an international conference on computing education research*, (New York, NY: ACM), 165–169. doi: 10.1145/2674683.2674699

Nguyen, J. G., Keuseman, K. J., and Humston, J. J. (2020). Minimise online cheating for online assessments during the COVID-19 pandemic. *J. Chem. Educ.* 97, 3429–3435. doi: 10.1021/acs.jchemed.0c00790

Rahim, A. F. A. (2020). Guidelines for online assessment in emergency remote teaching during the COVID-19 pandemic. *Educ. Med. J.* 12, 59–68. doi: 10.21315/eimj2020.12.2.6

Schneider, S. L., and Council, M. L. (2020). Distance learning in the era of COVID-19. *Arch. Dermatol. Res.* 313, 389–390. doi: 10.1007/s00403-020-02088-9

Scull, J., Phillips, M., Sharma, U., and Garnier, K. (2020). Innovations in teacher education at the time of COVID19: An Australian perspective. *J. Educ. Teach.* 46, 497–506. doi: 10.1080/02607476.2020.1802701

Sharadgah, T., and Sa'di, R. (2020). Preparedness of institutions of higher education for assessment in virtual learning environments during the Covid-19 lockdown: Evidence of bona fide challenges and pragmatic solutions. *J. Inf. Technol. Educ. Res.* 19, 755–774. doi: 10.28945/4615

Simpson, O. (2013). *Supporting Students in Online, Open and Distance Learning*, (3rd Edn). London: RoutledgeFalmer, doi: 10.4324/9780203417003

Tam, A. C. F. (2022). Students' perceptions of and learning practices in online timed take-home examinations during Covid-19. *Assess. Eval. High. Educ.* 47, 477–492. doi: 10.1080/02602938.2021.1928599

Van der Mescht, H. (2004). Phenomenology in education: A case study in educational leadership. *Indo Pac. J. Phenomenol.* 4, 1–16. doi: 10.1080/20797222.2004.11433887

Wang, Z. (2014). Developing Accuracy and Fluency in Spoken English of Chinese EFL Learners. *English Lang. Teach.* 7, 110–118.

White, A. (2020). *Amanda White on Education for and Detection of Contract Cheating in Virus Times. Campus Morning Mail*. Available online at: https://campusmorningmail.com.au/news/education-forand-detection-of-contract-cheating-in-the-age-of-covid-19/ (accessed on 15 Mar, 2021).

Wigfield, A., and Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Dev. Rev.* 12, 265–310. doi: 10.1016/0273-2297(92)90011-P

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

Yan-Li, S., Jiang, N., Pamanee, K., and Sriyanto, J. (2022). Online learning readiness and satisfaction during covid-19 pandemic among university students in four Asian countries. *J. Nusantara Stud.* 7, 245–269. doi: 10.24200/jonus.vol7iss2pp245-269

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fpsyg.2022.996504

Check for updates

OPEN ACCESS

EDITED BY
Antonio P. Gutierrez de Blume,
Georgia Southern University,
United States

REVIEWED BY
Liquan Qi,
China Education Press Agency, China
Jianping Qin,
The Chinese Society of Education,
China

*CORRESPONDENCE
Yueqi Shi
syq_shi@126.com

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

RECEIVED 17 July 2022
ACCEPTED 29 August 2022
PUBLISHED 23 September 2022

CITATION
Shi Y and Qu S (2022) Analysis of the effect of cognitive ability on academic achievement: Moderating role of self-monitoring.
Front. Psychol. 13:996504.
doi: 10.3389/fpsyg.2022.996504

Analysis of the effect of cognitive ability on academic achievement: Moderating role of self-monitoring

Yueqi Shi* and Shaowei Qu

School of Humanities and Social Sciences, University of Science and Technology Beijing, Beijing, China

In this study, cognitive ability was classified into memory ability, representational ability, information processing ability, logical reasoning ability, and thinking conversion ability, and analyzed the effects of these five ability values on academic achievement. Structural equation modeling (SEM) was used to analyze the moderating effect of Self-monitoring between cognitive ability and Academic Achievement, using students' Self-monitoring as moderating variables. The results of the study showed that cognitive ability can significantly and positively affect academic achievement, while Self-monitoring can significantly moderate the effect of cognitive ability on academic performance, with a significant moderating effect on math subjects and English subjects among achievement subjects, and the higher the value of cognitive ability, the stronger the moderating effect.

KEYWORDS

cognitive ability, academic performance, moderating effect, structural equation modeling, self-monitoring

Introduction

In China's educational student evaluation system, colleges and universities usually classify students by their Academic Achievement. Liu (2019) found that Academic Achievement, especially in school-organized examinations, significantly affects the future development of students. Li (2019) found that Academic Achievement is a concentrated expression of student learning and practice, an expression of students' cognitive abilities, and the particulars of their learning profile and standard. In China's student growth evaluation system, academic achievement is almost the only reference for admission to colleges and universities, and is the most important goal of student learning. Many existing studies have identified the necessary role of cognitive abilities when it comes to student learning. Therefore, in teaching practice, teachers pay particular attention to the development and exploitation of students' cognitive abilities. However, other research has found that cognitive ability is not the single major determinant of students' academic achievement (Shao, 1983). Students' Academic Achievement may not depend exclusively on the level of cognitive ability, but is determined by a combination of their overall learning status.

However, little research has been done on this part of the study and on the mechanisms of the influence of cognitive ability on Academic Achievement.

This research analyzed the role of cognitive ability on students' academic achievement under the moderating effect of self-monitoring, using students' self-monitoring as moderating variables.

The single effect of cognitive ability

Cognitive ability refers to the human brain's ability to store memory, process and extraction of information, includes attention, memory and logical reasoning, and thinking transformation. This is a key mental quality of students' completion of learning activities (Sternberg and Sternberg, 2009), it is also among the most investigated and most reliable predictive factors of student academic achievement (Matthias et al., 2016; Vilia et al., 2017). Xu and Li (2015) found that concentration, working memory, and logical reasoning were significantly predictive factors of Chinese and math achievement in a research of 4,743 middle school students. Rohde and Thompson (2007) concluded that cognitive ability was a direct predictor of academic achievement, with a correlation of 0.38. Ian et al. (2006) found a correlation between cognitive ability and Academic Achievement of 0.81. In a study that followed over 70,000 UK students for 5 years. Paul performed a stepwise analysis by multivariate analysis to obtain normalized regression coefficients (β) to analyze the association between the dimension of logical reasoning and students' performance in science and chemistry over the three semesters, and found a comparable absolute relationship between reasoning ability and students' performance in science and chemistry (Grass et al., 2017). Liu measured the cognitive abilities of spatial imagery, computation, and information processing in 499 Chinese children and teamed up to correlate students' Academic Achievement in mathematics and Chinese over two consecutive school years and found considerable associations between visual-spatial imagery, computation, and information processing abilities and Academic Achievement (Liu et al., 2021). Most of the previous studies of this type have been on the single effect of cognitive ability on Academic Achievement at the individual student level (Kuncel et al., 2004; Miriam et al., 2011). Chen (2016) argues that cognitive ability has a significant impact on the future direction of students, students with strong cognitive ability more likely to attend general high school and those with weak cognitive ability only going to vocational school. In addition, the above findings support the knowledge process theory (Deary et al., 2006; Xu and Li, 2015), that is, the stronger the cognitive ability of students, They are able to extract key information more quickly and accurately and encode it accurately and efficiently in their memory, allowing the brain to output more and more effective information, resulting in greater academic achievement on exams (Liu and Wang, 2000; Zhang and Zhang, 2011). Conversely, at lower levels of cognitive ability, some knowledge would be missed during the knowledge process, which would further reduce the effective information output and lead to lower academic achievement (Miriam et al., 2011). These findings also support the results of the previous analyses, suggesting that cognitive abilities typically contribute significantly to academic achievement.

While the correlation is clear between students' cognitive ability and academic achievement, it remains very challenging to clarify the complex mechanisms that influence them.

Through the research of many scholars, we can find that the association between students' cognitive abilities and academic achievement is clear, but the mechanisms of their complex influence remain very ambiguous. The cognitive abilities are important in students' study activities only through the researcher's predetermined scope of investigation that includes particular cognitive abilities, but outside the survey, these cognitive abilities are still operating in unpredictable ways (David, 2005), thus, the scholars still have not reached a consensus on the reasons why cognitive ability affects academic achievement due to the inconsistency in investigating the range of students' cognitive abilities (Formazin et al., 2011). The study by Xu and Li (2015) concluded that there was a significant correlation between attention and academic achievement, with the correlation coefficient ranging between 0.4 and 0.5. However, Zhang (2008) found that logical reasoning ability (LRA) had a correlation coefficient of around 0.3 with both Chinese and math scores, while attention was not significantly related to either subject. These findings all support the fact that cognitive ability can influence academic achievement indirectly, but this still does not lead to accurate conclusions about the advanced effect of individual student factor on academic achievement.

The moderating role of self-monitoring

Many previous studies have shown that there is indeed a link between the academic achievement of students with the personality factors of them. Gerhard (1996) concluded from an empirical analysis and concluded that personality traits contribute significantly to the academic achievement of students. Leino (1989) considered that non-intelligent factors and personality traits are the major causes of academic achievement. Gough (1964) studied 18 personality factors in secondary school students using the California Psychological Inventory (CPI) and correlated them with Academic Achievement and found that at least eight personality factors (e.g., control, responsibility, tolerance, and independence) were significantly related to Academic Achievement. Richardson et al. (2009) found that both rigor and achievement motivation among personality traits were significantly correlated with students' GPA by measuring five major personality traits and achievement motivation, and that this effect was moderated by achievement motivation. Adrian and Tomas (2005) analyzed the correlation between students' personality traits and their level of knowledge and showed that the rigorous and open personality traits were significantly correlated with the level of knowledge.

Self-monitoring, as a behavioral manifestation external to personality traits, also directly affects academic achievement. Callan and Cleary (2019) found that self-monitoring showed a significant positive correlation with math achievement and was a unique predictor of student achievement through an analysis of 96 8th grade students. Sabry (2010), through a follow-up study of 119 students found that self-management and self-monitoring significantly predicted students' academic achievement. Sawhney and Bansal (2015) found significant differences in academic achievement between students with strong and weak self-monitoring through a study of 100 students.

On the correlation between cognitive abilities and self-monitoring, many scholars have done research, but it is mainly biased toward qualitative research or analysis of the correlation between the two, so the interaction relationship between them has been little studied. Cognitive ability and self-monitoring are measured using a variety of dimensions, so the analysis of their correlations has different conclusions (James et al., 2006). Cognitive ability and self-monitoring are both separate and interrelated elements of individual student psychology (Li and Zhang, 2015).

Most self-monitoring are minimally correlated with cognitive ability, and therefore, cognitive ability and self-monitoring are often used as independent variables that affect students' academic achievement (Roemer et al., 2022); However, some scholars have argued that cognitive ability and self-monitoring interact (Borghans et al., 2008), Wang (2004) found through a study that self-monitoring in metacognition enhances students' learning ability and compensates for learning difficulties due to uneven cognitive development. Hayat and Shateri (2019) found through a study of 225 students that learning ability has a direct and positive effect on self-monitoring in metacognitive strategies. Zhang and Wei (2012) found that good self-monitoring strategies help to enhance students' learning ability by examining the relationship between students' self-monitoring strategies and their self-directed learning ability.

The inconsistent findings of the above studies may be related to the fact that Self-monitoring play more of a moderating role in Academic Achievement. Among the effects of cognitive ability on academic performance, the positive moderating effect of self-monitoring was found to be significant (Zhu et al., 2017), and that higher indices of self-regulation in Self-monitoring enable students to focus more on tasks and achieve better Academic Achievement (Nesayan et al., 2019), and self-regulation also significantly enhances the positive effect of cognitive ability on academic achievement (Shi and Qu, 2021).

Research hypothesis

In this study, cognitive abilities were classified into memory ability (MA), information processing ability (IPA), logical reasoning ability (LRA), representational ability (RA), and thinking conversion ability (TCA) based on the classification of cognitive abilities by Wo and Lin (2000) and Liang et al. (2020). Cognitive abilities will be analyzed for their role in influencing academic achievement and the following hypotheses will be formulated.

> **Hypothesis 1:** Cognitive abilities can influence academic achievement positively.

Meanwhile, the available research still analyses the effects of cognitive ability and self-monitoring as separate factors on students' academic achievement, while the effect of both together on academic achievement is unclear. Regarding the relationship between cognitive ability and self-monitoring on academic achievement, several empirical studies have shown that self-monitoring can significantly moderate the relationship between the influence of cognitive ability on students' knowledge, Freeman et al. (2017) found by examining the relationship between working memory capacity, self-monitoring, and academic achievement in 73 4th grade students that self-monitoring can influence the performance of working memory capacity and is reflected in and in academic achievement. Alberto et al. (2021) found in a study of 133 students aged 6–9 years that working memory capacity was not significantly different in its effect on overall achievement and grades in language and mathematics subjects. Leue et al. (2014) investigated the relationship between self-monitoring and the cognitive abilities of working memory, logical reasoning, language, and representational skills. The results found that self-monitoring and the moderation of working memory capacity were the most prominent, and self-monitoring also had significant moderation effects on logical reasoning capacity and representational capacity. This research hypothesizes that self-monitoring can also significantly and positively moderate the influence of cognitive ability on academic achievement and proposes the following hypotheses.

> **Hypothesis 2:** A positive moderating effect of self-monitoring between cognitive ability and academic achievement.

Materials and methods

Participants

The ethical examination of this study was approved by the Research Ethics Committee of the School of Humanities and Social Sciences, University of Science and Technology Beijing, and the study was conducted in accordance with the regulations for the protection of human subjects. This study selected 569 students as samples, all aged 15–18 years old, As shown in Table 1.

Procedure

Both the cognitive ability and personality trait measures in this research were conducted on campus. The students who took the test were organized by staff and tested in a separate classroom. The entire test lasted for 2 h.

Structural equation models were developed based on the cognitive ability, self-monitoring ability and academic achievement of the students obtained, and comprehensive achievement models and sub-subject models for Chinese, mathematics and English were developed, respectively. Each model was first tested for common method deviation in the analysis process, and then the fit of the model was tested according to the CFA test process. Finally, the influence of cognitive ability on academic achievement and the significance of the moderating effect of self-monitoring were analyzed according to the structural equation model, and the moderating effect was specifically analyzed using a simple slope test.

Measures

Cognitive ability

A stimulus-informed cognitive ability system designed by Wo (2010) was used for the cognitive ability test. The test uses computerized testing methods with techniques such as subtractive response times and additive response times (accurate to nanoseconds). The total number of correct fixed points of student feedback and the fixed feedback time were recorded throughout the testing process, and the accuracy of students' cognitive ability was analyzed based on the feedback records. The cognitive accuracy of the tested students was obtained by statistical methods, and their corresponding cognitive ability values were quantified and the quantified values were converted into T-scores to obtain the final cognitive ability values of the tested students. The final cognitive ability values contain five ability values: MA, LRA, RA, IPA, and TCA. The test method has been patented as an invention, and the sample size of the general test exceeds 2 million. The values of students' cognitive abilities obtained from the test were normally distributed with a range of trends of ±50 centered at 100, with high discriminant validity. The Cronbach's alpha of the test ranged among 0.80–0.90.

TABLE 1 Distribution of participating students.

Grade	Number of students			
	Boys	Proportion	Girls	Proportion
First grade	114	50.89%	110	49.11%
Second grade	82	46.33%	95	53.67%
Third grade	92	54.76%	76	45.24%
Total	288	50.62%	281	49.38%

Self-monitoring

The Student Personality Trait Scale was designed by Zhang et al. (2005). The scale uses a 5-point Likert scale: 5 (very much the same), 4 (comparatively the same), 3 (uncertain), 2 (relatively different), and1 (very different). There were forty-eight questions, consisting of four dimensions: planning, self-control, persistence and daring, with 12 questions set for each dimension. After the student assessment, the student's corresponding Self-monitoring dimension value was obtained by accumulating the scores of each dimension, and converted to T-score as a value of the students' personality trait dimensional ability. The Cronbach alpha coefficient for all the dimensions ranged from 0.60 to 0.93, with a validity of 0.91 and a test-regression reliability of 0.85.

Academic achievement

In the current research, in order to reduce the influence due to the level of students' test performance, the average of the students' four test scores in the semester when the cognitive ability was tested to be worthy was used as the academic score for each subject, and the raw scores were standardized (scores were assigned according to levels, with the highest score being 100 and the lowest being 0). In this study, three subjects, Chinese, Math and English, were selected for the study, and the composite academic score was the sum of the three subject's total scores.

Data analysis

This research first analyzed the correlation between cognitive ability, self-monitoring, and academic achievement, and then used structural equation modeling to analyze the moderating role of self-monitoring based on the moderating utility modeling procedure suggested by Wen and Ye (2014), and analyzed the pattern of the moderating role through a simple slope test. SPSS 25.0 and Mplus8.3 software were used to analyze the data.

Results

Common method deviation test

This study used a questionnaire in which the dimensions and sequence of test questions were set randomly, which reduced the bias introduced by self-designed questionnaires. Harman's single-factor test was used to test the effects of the procedure (Podsakoff et al., 2003), and an exploratory factor analysis of two variables (cognitive ability, and self-monitoring) was conducted. The results showed that the characteristic roots of all nine factors were higher than 1 after factor rotation, with the variance explained by the first factor being 33.38% (less than 40% of the critical value), so the variation level of the used methods is within an acceptable level (Wang et al., 2016).

Descriptive and bivariate analyses

The means, standard deviations, and correlation coefficients of each study variable are shown in **Table 2**. Cognitive ability, self-monitoring, and academic performance all showed significant positive correlations. For more data see **Supplementary Table**.

Measurement model check

An exploratory factor analysis was required to check the quality of the model before conducting the moderating effects analysis. Two latent variables were included for the present research, which were cognitive ability (five indicators including MA, IPA, LRA, RA, and TCA) and self-monitoring (four indicators including planning, self-control, persist, and daring). The test results showed that the model fit was good. $\chi^2(26) = 88.084$, CFI = 0.973, TLI = 0.962, SRMR = 0.042, RMESA = 0.065, and the 90% confidence interval for RMSEA was [0.050, 0.080], which indicated that the fitted indicators were within the normal reception range. It is also shown in **Table 3** that all of the latent variable indicators had significant standardized loadings on the corresponding factors ($p < 0.001$).

Moderating model checking

This study used structural equation modeling for examining the influence of cognitive ability, self-monitoring, and interaction terms on academic achievement, and the results are shown in **Figure 1**. On the basis of the moderated process of analysis suggested by Wen and Ye (2014), the impact of cognitive ability on academic achievement was analyzed using structural equation modeling. The results showed that cognitive ability significantly and positively influenced academic

TABLE 2 Means, standard deviations, and intercorrelations for variables.

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1. MA	106.58	13.39	1												
2. IPA	105.17	11.396	0.477**	1											
3. RA	107.9	6.812	0.431**	0.464**	1										
4. LRA	106.12	7.937	0.381**	0.429**	0.301**	1									
5. TCA	97.35	14.698	0.483**	0.592**	0.514**	0.483**	1								
6. PLANING	103.37	14.72	0.316**	0.420**	0.163**	0.310**	0.359**	1							
7. SELFCONTROL	107.17	14.27	0.423**	0.485**	0.219**	0.360**	0.423**	0.594**	1						
8. PERSISTENCE	104.66	14.893	0.429**	0.465**	0.192**	0.336**	0.413**	0.674**	0.512**	1					
9. DARING	101.66	15.638	0.430**	0.461**	0.270**	0.346**	0.432**	0.564**	0.609**	0.648**	1				
10. CHINESE	57.81797666	25.72765725	0.449**	0.496**	0.322**	0.390**	0.496**	0.530**	0.545**	0.574**	0.555**	1			
11. MATHEMATICS	50.47993002	28.91223569	0.426**	0.530**	0.319**	0.376**	0.481**	0.465**	0.530**	0.520**	0.502**	0.275**	1		
12. ENGLISH	52.20492511	28.08333079	0.444**	0.396**	0.202**	0.380**	0.397**	0.495**	0.526**	0.528**	0.519**	0.299**	0.117**	1	
13. TS	160.5028318	58.07603715	0.631**	0.681**	0.402**	0.548**	0.657**	0.612**	0.666**	0.675**	0.653**	0.730**	0.683**	0.679**	1

$N = 569$, $^*p < 0.05$, $^{**}p < 0.001$.

TABLE 3 Factor loading coefficient table.

Variable	Non-std. (Coef.)	SD	*z* (CR)	Std.
Cognitive ability				
MA	1	–	0.833301027	0.650***
IPA	1.004	0.069		0.766***
RA	0.462	0.039		0.591***
LRA	0.532	0.045		0.583***
TCA	1.313	0.091		0.778***
Personality traits				
PLANING	1	–	0.886514422	0.743***
SELF-CONTROL	1.06	0.056		0.812***
PERSIT	1.186	0.058		0.871***
DARE	1.089	0.061		0.762***

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

achievement ($\beta = 0.868$, $p < 0.001$). Then, self-monitoring and interaction terms were brought into the model as moderating variables.

Model 1: Impact on total academic achievement

The values of the goodness-of-fit indicators of the structural equation model were $\chi^2(72) = 332.136$, CFI = 0.945, TLI = 0.930, SRMR = 0.073, RMESA = 0.079, the 90% confidence interval of RMSEA is [0.071, 0.088], which shows that the values of the goodness-of-fit indicators of the model were all in the acceptable range and the fit was good. The result is shown in Figure 1.

As seen in Figure 1, cognitive ability has positive and significant correlation with total academic achievement ($\gamma = 0.695$, $p < 0.001$), Self-monitoring has positive and significant correlation with total academic achievement ($\gamma = 0.397$, $p < 0.001$), and the interaction term also has positive and significant correlation with total academic achievement ($\gamma = 0.323$, $p < 0.005$), indicating that the Self-monitoring between cognitive ability and total Academic Achievement moderating effect is significant. Therefore, hypothesis1,2 is valid.

To reveal more clearly the specific pattern of moderating effects, a simple slope test was further used to analyze the moderating role of Self-monitoring. Firstly, subjects were divided into high support group ($Z \geq 1SD$), low support group ($Z \leq 1SD$), and medium support group (between the above two groups) according to the mean score of Self-monitoring plus or minus one standard deviation; secondly, group regression was used to examine the relationship between cognitive ability and total Academic Achievement, controlling for variables such as gender and grade. The results showed that cognitive ability in all three personality trait groups significantly predicted total Academic Achievement ($p < 0.001$), with a prediction coefficient β of 0.372 for the low Self-monitoring group, 0.695 for the medium Self-monitoring group, and 1.018 for the high personality trait group. The results are shown in Figure 2. As the value of Self-monitoring increases, the prediction of cognitive ability on total Academic Achievement tends to increase gradually; and as the cognitive ability increases, the higher the value of students' Self-monitoring, the higher the total Academic Achievement.

Model 2: Impact on Chinese academic achievement

The values of the goodness-of-fit indicators of the structural equation model were $\chi^2(72) = 468.731$, CFI = 0.864, TLI = 0.829, SRMR = 0.100, RMESA = 0.098, the 90% confidence interval of RMSEA is [0.090, 0. 107], which shows that the values of the goodness-of-fit indicators of the model were relatively general. The result is shown in Figure 3.

As seen in Figure 3, cognitive ability has positive and significant correlation with Chinese academic achievement ($\gamma = 0.616$, $p < 0.001$), Self-monitoring has positive and significant correlation with Chinese academic achievement ($\gamma = 0.471$, $p < 0.001$). But the interaction term could not positively predict Chinese Academic Achievement ($\gamma = 0.430$, $p > 0.01$), indicating that the Self-monitoring between cognitive ability and Chinese Academic Achievement moderating effect is not significant.

Model 3: Impact on mathematics academic achievement

The values of the goodness-of-fit indicators of the structural equation model were $\chi^2(72) = 337.206$, CFI = 0.977, TLI = 0.969, SRMR = 0.039, RMESA = 0.056, the 90% confidence interval of RMSEA is [0.043, 0.070], which shows that the values of the goodness-of-fit indicators of the model were all in the acceptable range and the fit was good. The result is shown in Figure 4.

As seen in Figure 4, cognitive ability has positive and significant correlation with mathematics academic achievement ($\gamma = 0.458$, $p < 0.001$), Self-monitoring has positive and significant correlation with mathematics academic achievement ($\gamma = 0.293$, $p < 0.001$), and the interaction term also has positive and significant correlation with mathematics academic achievement ($\gamma = 0.162$, $p < 0.005$), indicating that the Self-monitoring between cognitive ability and mathematics Academic Achievement moderating effect is significant.

To reveal more clearly the specific pattern of moderating effects, a simple slope test was further used to analyze the moderating role of Self-monitoring. Firstly, subjects were divided into high support group ($Z \geq 1SD$), low support group ($Z \leq 1SD$), and medium support group (between the above two groups) according to the mean score of Self-monitoring plus or minus one standard deviation; secondly, group regression was used to examine the relationship between cognitive ability and mathematics Academic Achievement, controlling for variables such as gender and grade. The results showed that cognitive ability

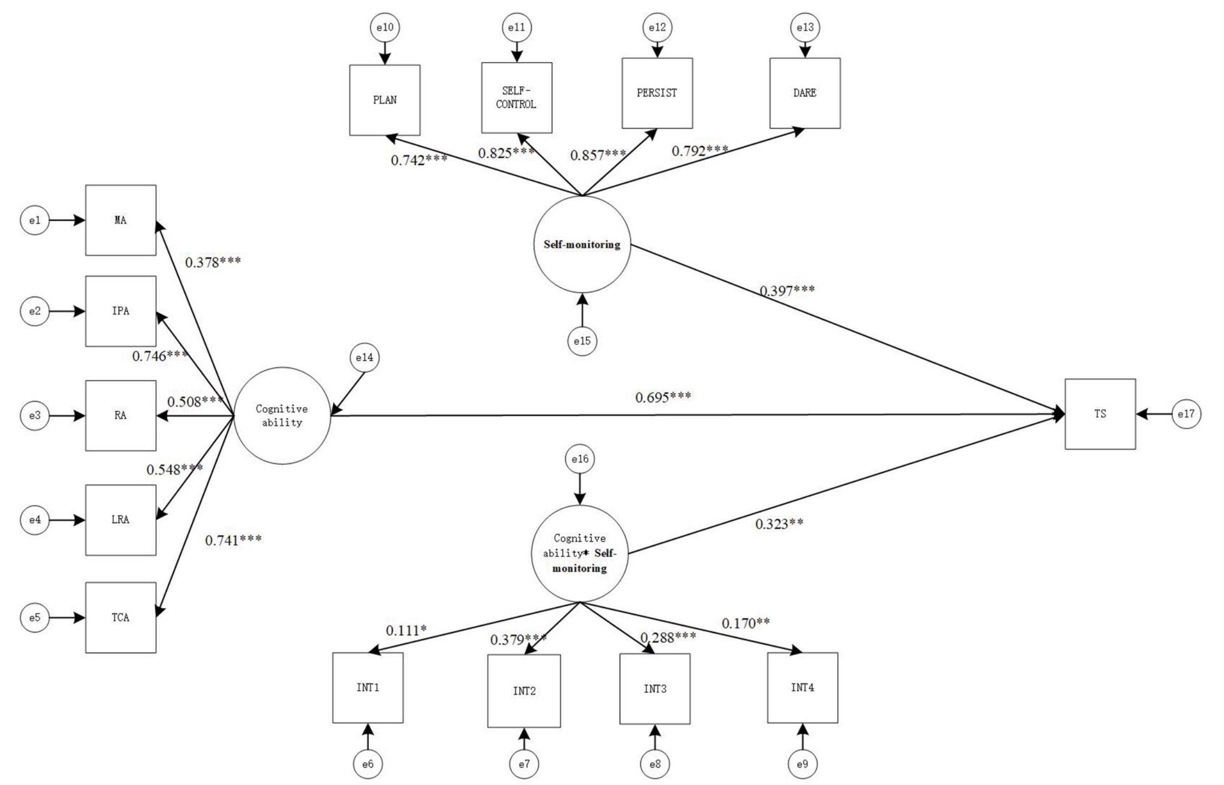


FIGURE 1
Structural equation moderating effect relationship model diagram (model 1). MA, memory ability; IPA, information processing ability, RA, representation ability; LRA, logical reasoning ability; TCA, thinking conversion ability; TS- academic performance. *p < 0.05, **p < 0.01, and ***p < 0.001.

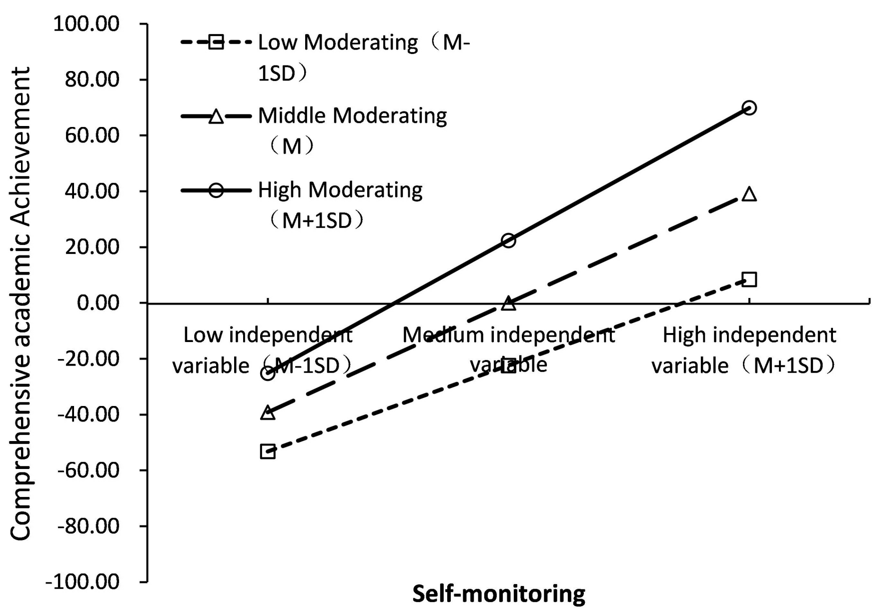


FIGURE 2
Simple slope test (model 1).

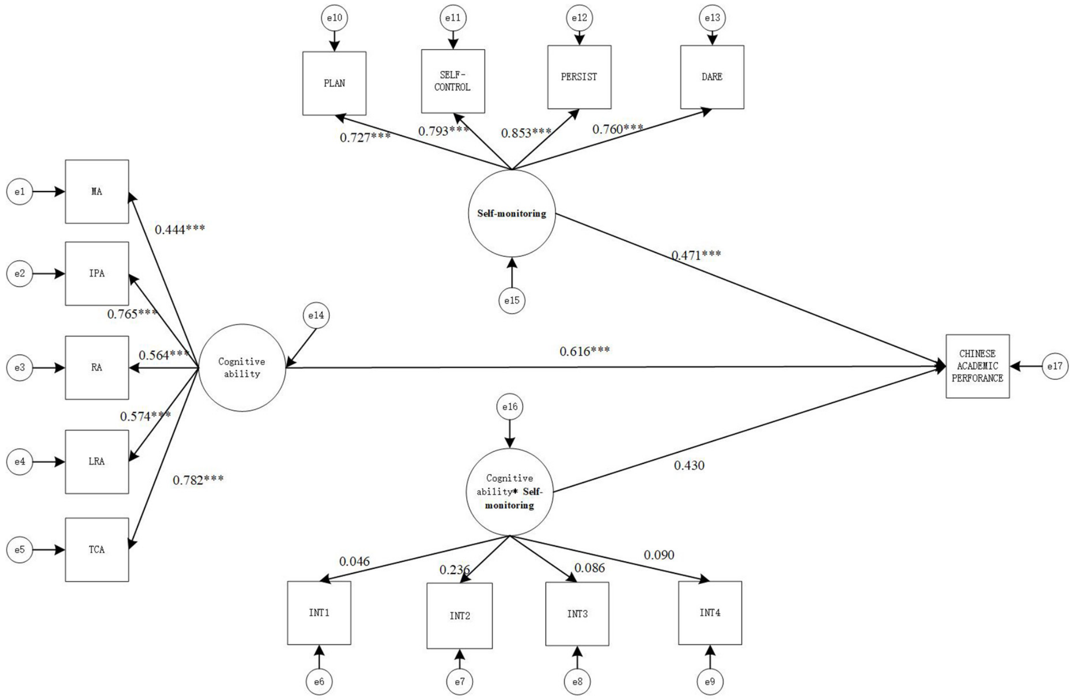


FIGURE 3
Structural equation moderating effect relationship model diagram (model 2). MA, memory ability; IPA, information processing ability; RA, representation ability; LRA, logical reasoning ability; TCA, thinking conversion ability.*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

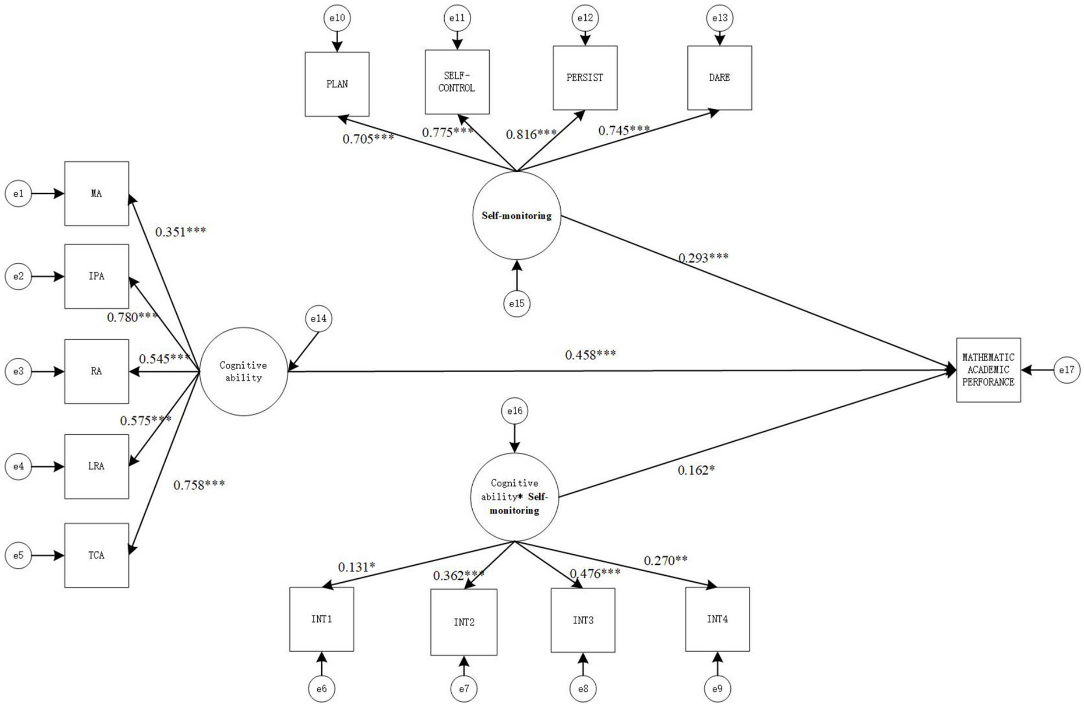


FIGURE 4
Structural equation moderating effect relationship model diagram (model 3). MA, memory ability; IPA, information processing ability; RA, representation ability; LRA, logical reasoning ability; TCA, thinking conversion ability.*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

in all three personality trait groups significantly predicted mathematics Academic Achievement ($p < 0.001$), with a prediction coefficient β of 0.296 for the low Self-monitoring group, 0.458 for the medium Self-monitoring group, and 0.620 for the high personality trait group. The results are shown in **Figure 5**. As the value of Self-monitoring increases, the prediction of cognitive ability on mathematics Academic Achievement tends to increase gradually; and as the cognitive ability increases, the higher the value of students' Self-monitoring, the higher the mathematics Academic Achievement.

Model 4: Impact on English academic achievement

The values of the goodness-of-fit indicators of the structural equation model were $\chi^2(72) = 384.832$, CFI = 0.970, TLI = 0.960, SRMR = 0.042, RMESA = 0.064, the 90% confidence interval of RMSEA is [0.051, 0.077], which shows that the values of the goodness-of-fit indicators of the model were all in the acceptable range and the fit was good. The result is shown in **Figure 6**.

As seen in **Figure 6**, cognitive ability has positive and significant correlation with English academic achievement ($\gamma = 0.630$, $p < 0.001$), Self-monitoring has positive and significant correlation with English academic achievement ($\gamma = 0.377$, $p < 0.001$), and the interaction term also has positive and significant correlation with English academic achievement ($\gamma = 0.192$, $p < 0.005$), indicating that the Self-monitoring between cognitive ability and English Academic Achievement moderating effect is significant.

To reveal more clearly the specific pattern of moderating effects, a simple slope test was further used to analyze the moderating role of Self-monitoring. Firstly, subjects were divided into high support group ($Z \geq 1SD$), low support group ($Z \leq 1SD$), and medium support group (between the above two groups) according to the mean score of Self-monitoring plus or minus one standard deviation; secondly, group regression was used to examine the relationship between cognitive ability and English Academic Achievement, controlling for variables such as gender and grade. The results showed that cognitive ability in all three personality trait groups significantly predicted English Academic Achievement ($p < 0.001$), with a prediction coefficient β of 0.438 for the low Self-monitoring group, 0.630 for the medium Self-monitoring group, and 0.822 for the high personality trait group. The results are shown in **Figure 7**. As the value of Self-monitoring increases, the prediction of cognitive ability on English Academic Achievement tends to increase gradually; and as the cognitive ability increases, the higher the value of students' Self-monitoring, the higher the English Academic Achievement.

Discussion

The single effect of cognitive ability on academic achievement

Many researches have obtained that cognitive ability is a very important pyschological characteristic of students and that

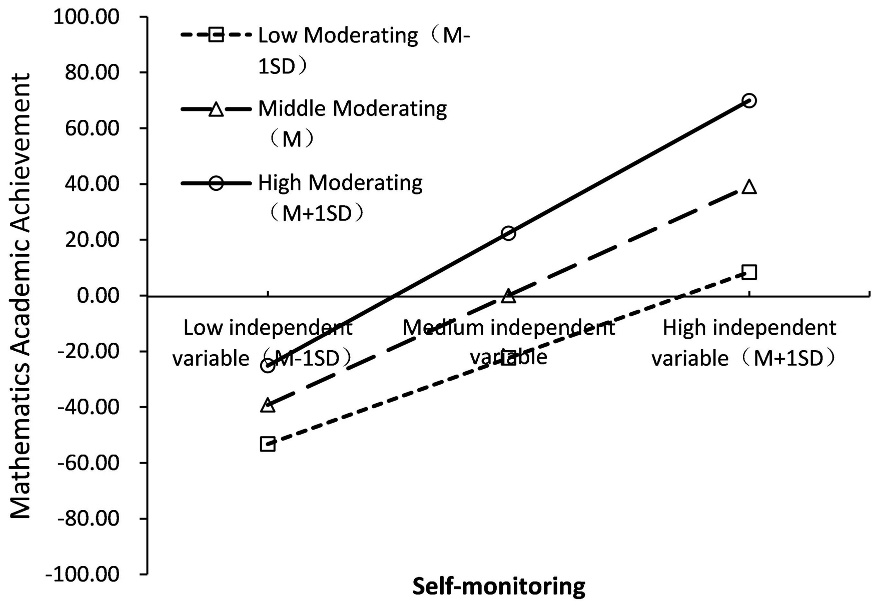


FIGURE 5
Simple slope test (model 3).

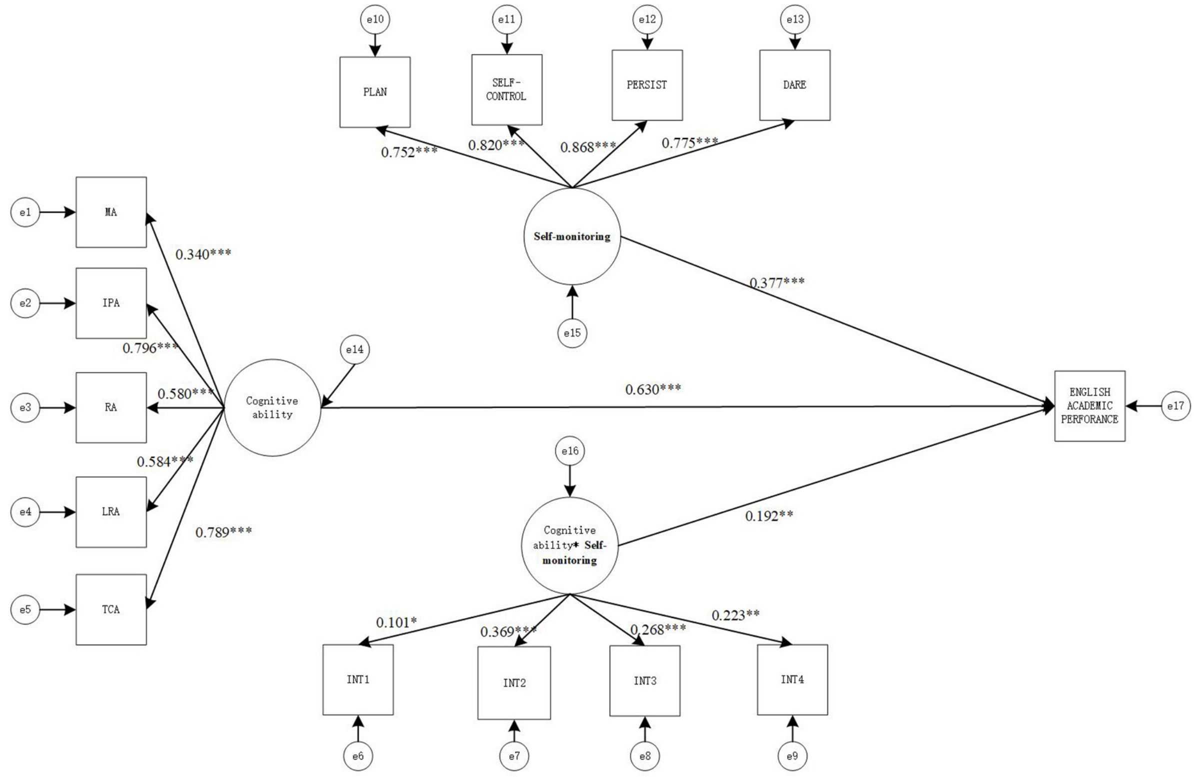


FIGURE 6
Structural equation moderating effect relationship model diagram (model 4). MA, memory ability; IPA, information processing ability; RA, representation ability; LRA, logical reasoning ability; TCA, thinking conversion ability. $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

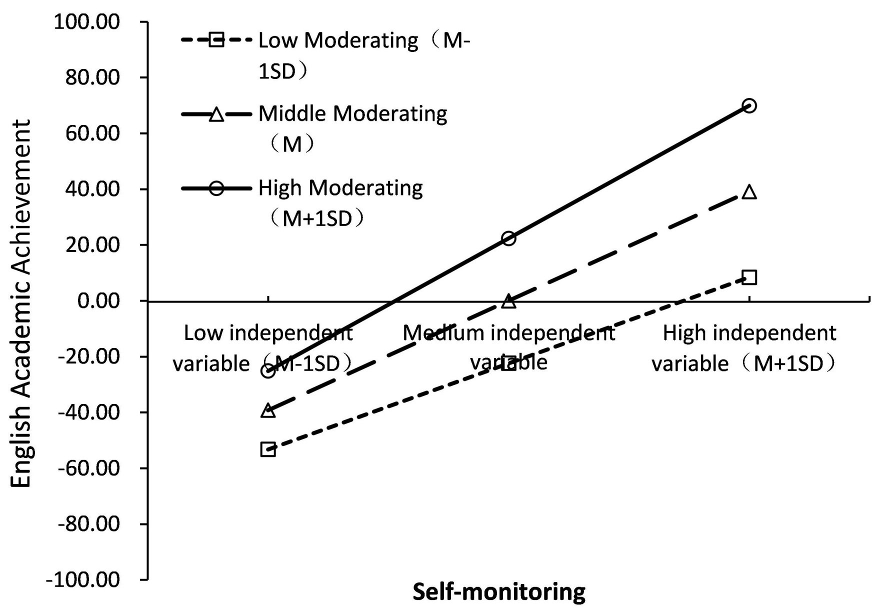


FIGURE 7
Simple slope test (model 4).

cognitive ability ensures the successful achievement of learning activities (Matthias et al., 2016).

IPA (information processing ability) is mainly the ability of students to understand information after acquiring it through reading and listening, which is closely related to students' classroom efficiency. Students with better IPA ability can fully understand and master the content taught by the teacher in class and quickly construct it into their own knowledge system, thus enhancing their knowledge mastery and achieving better Academic Achievement in exams (Yuan and Wen, 2003).

TCA (Thinking Conversion Ability) is the active situation of students' thinking in understanding knowledge during the learning process, which is mainly reflected in the rate and accurately of thinking conversion, therefore, no subject can be learned without the ability. In mathematics, in particular, students with strong TCA ability are more likely to summary and acquire thinking and skills to solve new mathematical questions, and to make analogies in similar problems, thus improving the accuracy of solving novel and difficult problems, and thus improving their academic performance (Liu, 1988).

MA (Memory Ability) is mainly a student's long term memory ability. The stronger the MA ability value, the more students are able to remember knowledge quickly and keep it for a long time. In addition, MA can interact with IPA to significantly improve students' reading ability by mastering more reading information through memory ability and thus comprehending information more quickly during reading, which is particularly evident in reading comprehension questions on both Chinese and English exams. As a result, Academic Achievement is also better among students with strong MA proficiency values (Yu et al., 2014).

LRA (Logical Reasoning Ability) includes both types of inductive and deductive reasoning. LRA has a significant positive impact on mathematics academic achievement of students (Zhu et al., 2020). Recently, Chinese higher education entrance examinations have begun to focus on students' reasoning ability, which is also reflected in the Chinese and English subject examinations, where a large number of logical reasoning questions have been added (Hu, 2017); therefore, a significant positive effect of LRA has also been demonstrated on academic performance in both Chinese and English examinations.

RA (representational ability) helps students to understand three-dimensional spatial knowledge in mathematics subjects by constructing three-dimensional images and pictures in their own minds, which helps to understand spatial knowledge. In addition, RA can also stimulate associative memory during students' memorization process by forming images in their minds while they recite knowledge related to Chinese and English, which leads to stronger and longer-lasting memory for knowledge and thus better Academic Achievement (Lin et al., 2003).

Meanwhile, we can find that the effect of MA on Academic Achievement in Chinese and English (correlation coefficients of 0.439 and 0.435, respectively) is greater than that on Academic Achievement in mathematics (correlation coefficient of 0.424), while the effect of IPA on Academic Achievement in mathematics (correlation coefficient of 0.531) is much greater than that on Academic Achievement in Chinese (correlation coefficient of 0.496) and English (correlation coefficient of 0.393). It can be surmised that for the subjects of Chinese and English, the test focuses on students' memorized knowledge, while for the subject of mathematics, the test focuses on students' knowledge processing and information extraction skills.

The moderating role of self-monitoring

Significant positive moderating effects of self-monitoring between cognitive ability and academic performance were observed, and the higher the value of students' cognitive ability, the more pronounced the effect of the moderating component.

Students' behavior in learning activities is guided by their individual stable self-monitoring by nature, may receive indirect influence from the learning environment, but is essentially determined by their own self-monitoring (Nie et al., 2011). During the learning process, the individual student's typical response to environmental stimuli is real-time, emotional, and conditioned. However, individual students can prevent the triggering of external stimuli through specific manifestations of self-monitoring so that they do not react impulsively and in a timely manner. Self-monitoring can influence and temporarily alter individual behavioral characteristics by organizational strategies including planning, persistence, self-control, and daring, thereby reducing impulsive behavior and producing other positive learning behaviors in response to situations (Mischel and Mischel, 1983; Mischel and Shoda, 1998).

Usually, those students who have higher levels of self-planning, self-restraint, self-persistence and self-regulation within their self-monitoring are able to keep more stable information emotionally during the learning process in the face of learning pressure from the external environment, thus ensuring that their learning efficiency is not affected by the outside world.

In self-determination theory, good academic behavior outcomes (e.g., Academic Achievement) can occur if students meet specific requirements for psychological trait stability (e.g., persistence and self-discipline) (Ryan and Deci, 2000). In this way, there is a virtuous cycle of learning that students can establish, including planning, self-regulated execution, persistence, and dynamic adjustment, so that they can continuously improve their learning process and achieve better Academic Achievement in examinations (Lin, 2020).

Students with high planning and self-control in their Self-monitoring tend to have more time to study and higher learning efficiency, which allows students to fully utilize their cognitive abilities, thus allowing them to achieve academic satisfaction with less time spent. In addition, the persistence and dare nature of personality characteristic are very helpful for students to challenge learning difficulties, and together with strong cognitive ability, students can easily solve learning problems, thus gaining a greater sense of academic achievement and higher Academic Achievement.

Usually, students with high learning efficiency will also have stronger learning self-confidence; therefore, when students have strong values of Self-monitoring, their learning efficiency will also be more prominent, which is more obvious when students have stronger cognitive ability, the stronger students' cognitive ability, the higher students' learning self-confidence, the more willing students are to take the initiative to learn, and thus their learning efficiency will be improved, which in turn will improve their Academic Achievement (Liang et al., 2020). Therefore, as a result, it is clear that Self-monitoring can significantly moderate the effect of cognitive ability on academic performance, and the stronger the cognitive ability, the more prominent the moderating effect.

In addition, in the analysis of the moderating effect of Chinese, mathematics, and English subjects, it can be found that the moderating effect of Self-monitoring is significant for mathematics and English subjects, but not for Chinese subjects. This is because the improvement of Chinese subjects often requires students to invest large amounts of time and effort to accumulate, which is very related to students' interest and not much related to the strength of cognitive ability; while Mathematics and English are subjects that can be improved if students follow the plan carefully. Therefore, the Self-monitoring of planning, self-control, persistence and daring are very helpful in helping students to learn according to the teacher's requirements, and the stronger the cognitive ability, the more effective the learning will be, resulting in higher Academic Achievement.

The simple slope test reveals that self-monitoring is a consistent mechanism in moderating overall academic performance, math performance, and English performance. It is that the moderating effect is significantly greater in high self-monitoring ability than in low self-monitoring ability. Students with high levels of self-monitoring have more knowledge about learning and learning strategies, and are good at controlling their own learning process, using various strategies to solve problems, and promoting the development of strong learning skills. Students with low levels of self-monitoring, on the other hand, lack knowledge about learning and learning strategies, and are not good at using different strategies according to the changes in materials and learning tasks, and usually show poor learning ability. Good self-monitoring skills enable individuals to acquire, organize and use information more effectively and avoid losses due to low cognitive ability. On the contrary, if the cognitive ability is high but the self-monitoring level is low, the lack of self-management and self-regulation of cognitive activities makes the cognitive activities lack planning and purpose, resulting in low efficiency of problem solving and lower possibility of success. The higher the level of self-monitoring, the better it is for improving learning efficiency, bringing into play the strengths of cognitive ability, and thus achieving better academic success.

Limitations and future directions

One of the most prominent limitations of this study is the small sample size and the single range of students surveyed. To further enhance the credibility of the study's findings, more schools in other Chinese provinces should be selected for the study and comparison. Furthermore, this study considered only the external effects caused by self-monitoring when analyzing the effects of cognitive ability on continued academic performance. As well, the effects of students' other psychological states on cognitive ability and Academic Achievement were not considered in the analysis. Future research could focus on this area to obtain more valuable findings.

Conclusion

This research used structural equation modeling to analyze the moderating effect of self-monitoring of cognitive ability on academic achievement. The results of the research showed that cognitive ability can positively influence academic achievement significantly, while Self-monitoring can significantly moderate the effect of cognitive ability on academic performance, with a significant moderating effect on math subjects and English subjects among achievement subjects, and the higher the value of cognitive ability, the stronger the moderating effect.

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 author.

Ethics statement

The studies involving human participants were reviewed and approved by the Research Ethics Committee of the School of Humanities and Social Sciences, University of Science and Technology Beijing.

Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

YS contributed to the conception and design of the study and performed the statistical analysis. YS and SQ contributed to the data collection and wrote the first draft of the manuscript. Both authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This research was received support from the National Natural Science Foundation of China (L2124028).

Acknowledgments

We thank the professional EDITAGE editing language services for ensuring that this manuscript is free of English grammatical errors.

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.2022.996504/full#supplementary-material

References

Adrian, F., and Tomas, C. P. (2005). Personality, intelligence and general knowledge. *Learn. Individ. Differ.* 16, 79–90. doi: 10.1016/j.lindif.2005.07.002

Alberto, Q. R., Nieves, M., and Alejandra, C. P. (2021). Task monitoring and working memory as executive components predictive of general and specific academic achievements in 6–9-year-old children. *Int. J. Environ. Res. Public Health.* 18:6681. doi: 10.3390/ijerph18136681

Borghans, L., Angela, L. D., James, J. H., and Bas, T. W. (2008). The economics and psychology of personality traits. *J. Hum. Resour.* 43, 972–1059. doi: 10.1111/j.0042-7092.2007.00700.x

Callan, G. L., and Cleary, T. J. (2019). Examining cyclical phase relations and predictive influences of self-regulated learning processes on mathematics task performance. *Metacogn. Learn.* 14, 43–63. doi: 10.1007/s11409-019-09191-x

Chen, W. (2016). The difference in income return between vocational education and ordinary high school education. *Chin. J. Sociol.* 2, 167–190.

David, F. L. (2005). The role of nonverbal ability tests in identifying academically gifted students: an aptitude perspective. *Gifted Child Q.* 49, 111–138. doi: 10.1177/001698620504900203

Deary, I. J., Steve, S., Pauline, S., and Cres, F. (2006). Intelligence and educational achievement. *Intelligence* 35, 13–21.

Formazin, M., Schroeders, U., Kller, O., Wilhelm, O., and Westmeyer, H. (2011). Studierendenauswahl im fach psychologie. *Psychol. Rundschau.* 62, 221–236. doi: 10.1026/0033-3042/a000093

Freeman, E. E., Karayanidis, F., and Chalmers, K. A. (2017). Metacognitive monitoring of working memory performance and its relationship to academic achievement in grade 4 children. *Learn. Individ. Diff.* 57, 58–64. doi: 10.1016/j.lindif.2017.06.003

Gerhard, B. (1996). Personality traits, learning stratigies, and performance. *Eur. J. Pers.* 10, 337–352. doi: 10.1002/(SICI)1099-0984(199612)10:53.0.CO;2-7

Gough, H. G. (1964). Academic achievement in middle school as predicted from the california psychological inventory. *J. Educ. Psychol.* 55:174. doi: 10.1037/H0046186

Grass, J., Strobel, A., and Strobel, A. (2017). Cognitive investments in academic success: the role of need for cognition at university. *Front. Psychol.* 8:790. doi: 10.3389/fpsyg.2017.00790

Hayat, A. A., and Shateri, K. (2019). The role of academic self-efficacy in improving students' metacognitive learning strategies. *J. Med. Edu. Profession* 7:205. doi: 10.30476/jamp.2019.81200

Hu, X. D. (2017). The principles and paths of the core competence of chinese in the college entrance examination. *J China Examinations* 07, 58–65. doi: 10.19360/j.cnki.11-3303/g4.2017.07.012

Ian, J. D., Steve, S., Pauline, S., and Cres, F. (2006). Intelligence and educational achievement. *Intelligence* 35, 13–21. doi: 10.1016/j.intell.2006.02.001

James, J. H., Jora, S., and Sergio, U. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *NBER Working Papers* 24, 411–482. doi: 10.1086/504455

Kuncel, N. R., Hezlett, S. A., and Ones, D. S. (2004). Academic achievement, career potential, creativity, and job performance: can one construct predict them all? *J. Pers. soc. Psychol.* 86, 148–161. doi: 10.1037/0022-3514.86.1.148

Leino, A. L. (1989). A study of learning styles. *Res. Bull.* 3, 56–68.

Leue, A., Weber, B., and Beauducel, A. (2014). How do working-memory-related demand, reasoning ability and aversive reinforcement modulate conflict monitoring? *Front. Hum. Neurosci.* 8:210. doi: 10.3389/fnhum.2014.00210

Li, D. Z. (2019). *The Influence of Parental Rearing Style and Self-Control on Autonomous Learning and Academic Achievement*. Wuhan: Huazhong Normal University.

Li, T., and Zhang, W. T. (2015). International trends in the study of personality economics. *Econ. Perspect.* 8, 128–143.

Liang, X. L., He, J., and Liu, P. P. (2020). The influence of cognitive ability on academic achievement of junior middle school students: a mediated moderation model. *Psychol. Dev Educ.* 449–461. doi: 10.16187/j.cnki.issn1001-4918.2020.04.08

Lin, C. D., Wo, J. Z., and Liu, H. J. (2003). An EEG study on the intelligence development of primary and middle school students. *Stud. Psychol. Behav.* 1, 5–10.

Lin, S. H. (2020). The relationship between initiative personality, study initiative and Academic Achievement of middle school students. *Mental Health Educ. Primary Second. Sch.* 18, 15–21.

Liu, J., and Wang, H. L. (2000). Application of information processing learning theory in educational technology. *Inserv. Educ. Train. Sch. Teach.* 12, 51–52.

Liu, S., Wei, W., Chen, Y., Hugo, P., and Zhao, J. (2021). Visual–Spatial Ability Predicts Academic Achievement Through Arithmetic and Reading Abilities. *Front. Psychol.* 11:591308. doi: 10.3389/fpsyg.2020.591308

Liu, X. J. (2019). *Analysis of Factors Affecting High School Students' College Entrance Examination Results*. Hunan: Hunan University.

Liu, Y. (1988). On the cultivation of thinking conversion ability in middle school mathematics teaching. *J Yangzhou Univ.* 110–114. doi: 10.19411/j.1007-824x.1988.03.021

Matthias, S., Miriam, A., Nicolas, B., and Samuel, G. (2016). Choosing between what you want now and what you want most: Self-control explains Academic Achievement beyond cognitive ability. *Pers. Individ. Differ.* 94, 168–172. doi: 10.1016/j.paid.2016.01.029

Miriam, V., Franzis, P., and Heinz, H. (2011). Mental abilities and school achievement: A test of a mediation hypothesis. *Intelligence* 39, 357–369. doi: 10.1016/j.intell.2011.06.006

Mischel, M. W., and Mischel, W. (1983). The development of children's knowledge of self-control strategies. *Child Dev.* 54, 603–619. doi: 10.2307/1130047

Mischel, W., and Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Ann. Rev. Psychol.* 49:229. doi: 10.1146/annurev.psych.49.1.229

Nesayan, A., Amani, M., and Asadi Gandomani, R. (2019). A study of the cognitive profile of children and its relationship with academic achievement. *Basic Clin. Neurosci. J.* 10, 165–174. doi: 10.32598/bcn.9.10.230

Nie, Y. G., Yang, A., and Zeng, M. X. (2011). Adolescents' metacognition, big five personality and the relationship between learning adaptation behavior. *Psychol. Dev. Educ.* 02, 151–157. doi: 10.16187/j.cnki.issn1001-4918.2011.02.011

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *J. Appl. Psychol.* 885, 10–1037. doi: 10.1037/0021-9010.88.5.879

Richardson, M., Abraham, C., Anderson, J. C., Gerbing, D. W., Bandalos, D. L., Finney, S. J., et al. (2009). Conscientiousness and achievement motivation predict performance. *Eur. J. Pers.* 23, 589–605. doi: 10.1002/per.732

Roemer, L., Lechner, C. M., and Rammstedt, B. (2022). Beyond competencies: associations between personality and school grades are largely independent of subject-specific and general cognitive competencies. *J. Intell.* 10:26. doi: 10.3390/jintelligence10020026

Rohde, T. E., and Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence* 35, 83–92. doi: 10.1016/j.intell.2006.05.004

Ryan, R. M., and Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 68–78.

Sabry, M. (2010). Garrison's model of self-directed learning: preliminary validation and relationship to academic achievement. *Spanish J. Psychol.* 13, 586–596. doi: 10.1017/S1138741600002262

Sawhney, N., and Bansal, S. (2015). Metacognitive awareness of undergraduate students in relation to their academic achievement. *Int. J. Indian Psychol.* 3, 107–114. doi: 10.25215/0301.136

Shao, R. Z. (1983). *Educational Psychology: Principles of Learning and Teaching*. Shanghai: Education Press.

Shi, Y. Q., and Qu, S. W. (2021). Cognitive ability and self-control's influence on high school students' total academic achievement. *Front. Psychol.* 12:783673. doi: 10.3389/fpsyg.2021.783673

Sternberg, R. J., and Sternberg, K. (2009). *Cognitive psychology*, 6th Edn. Belmont, CA: Wadsworth, Cengage Learning.

Vilia, P. N., Candeias, A. A., Neto, A. S., Franco, M. S., and Melo, M. (2017). Academic achievement in physics-chemistry: the predictive effect of attitudes and reasoning abilities. *Front. Psychol.* 8:1064. doi: 10.3389/fpsyg.2017.01064

Wang, Y. L., Qin, Y. L., Xiao, C., and Lin, X. Y. (2016). The relationship between parental conflict and adolescent self-harm: a moderated mediation model. *Psychol. Dev. Educ.* 3, 377–384. doi: 10.16187/j.cnki.issn1001-4918.2016.03.15

Wang, Y. N. (2004). Structure, function and development of metacognition. J. Nanjing Normal University. *Soc. Sci. Edition* 1:6. doi: :10.3969/j.issn.1001-4608-B.2004.01.015

Wen, Z. L., and Ye, B. J. (2014). Analysis of mediation effect: method and model development. *Adv. Psychol. Sci.* 731–745.

Wo, J. Z. (2010). *Stimulating Information Cognitive Ability Value Test System and Method*. Beijing: China National Intellectual Property Administration.

Wo, J. Z., and Lin, C. D. (2000). Research on the Development of Adolescents' Self-monitoring Ability. *J. Psychol. Sci.* 10–15+124. doi: 10.16719/j.cnki.1671-6981.2000.01.003

Xu, F., and Li, C. H. (2015). The influence of junior middle school students' cognitive ability on academic achievement. *J. Psychol. Sci.* 1, 11–18. doi: 10.16719/j.cnki.1671-6981.2015.01.033

Yu, B., Le, G. A., and Liu, H. J. (2014). Working memory and self-regulation. *Adv. Psychol. Sci.* 5, 772–781.

Yuan, L. X., and Wen, J. Y. (2003). Research on the relationship between middle school students' learning strategies and academic achievement. *J. Guangdong Univ. Educ.* 1, 7–12.

Zhang, H., Wo, J. Z., and Liu, C. M. (2005). "The formulation of the norm of personality scale for senior high school students," *Paper Presented at the 10th National Conference of Psychology*, ed. Chinese Psychological Society Shanghai: Chinese Psychological Society, 782.

Zhang, J. H., and Zhang, J. P. (2011). The evolution and thinking of information processing model in the learning process. *Educ. Res.* 1, 40–43.

Zhang, J. L. (2008). Comparison of logical reasoning ability, metacognition, attention level and Academic Achievement of primary and middle school students. *J. Psychol. Sci.* 707–710. doi: 10.16719/j.cnki.1671-6981.2008.03.035

Zhang, L., and Wei, Y. (2012). "A Study on Promoting the Autonomous Ability by Meta-cognitive Strategy in English Web-Based Environment," in *Frontiers in Computer Education*, eds S. Sambath and E. Zhu Berlin: Springer, 195–199. doi: : 10.1007/978-3-642-27552-4_29

Zhu, L. M., Ke, S., and Wu, L. S. (2020). Investigation and research on the development level of middle school students' mathematical and logical reasoning ability. *J. Tianjin Normal Univ.* 4, 18–23. doi: 10.16826/j.cnki.1009-7228.2020.04.004

Zhu, Y., He, W., and Wang, Y. (2017). Challenge–hindrance stress and academic achievement: Proactive personality as moderator. *Soc. Behav. Pers.* 45, 441–452. doi: 10.2224/sbp.5855

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 29 September 2022
DOI 10.3389/feduc.2022.1003378

Check for updates

OPEN ACCESS

EDITED BY
Antonio P. Gutierrez de Blume,
Georgia Southern University,
United States

REVIEWED BY
Hamid Sharif Nia,
Mazandaran University of Medical Sciences, Iran
Maryam Rassouli,
Shahid Beheshti University of Medical Sciences, Iran

*CORRESPONDENCE
Wilter C. Morales-García
wiltermorales@upeu.edu.pe

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

RECEIVED 26 July 2022
ACCEPTED 14 September 2022
PUBLISHED 29 September 2022

CITATION
Gonzales-Valdivia J, Morales-García WC, Saintila J, Huancahuire-Vega S, Morales-García M and Ruiz Mamani PG (2022) Translation and validation of the high-school satisfaction scale (H-SatP Scale) in Peruvian students. *Front. Educ.* 7:1003378.
doi: 10.3389/feduc.2022.1003378

Translation and validation of the high-school satisfaction scale (H-SatP Scale) in Peruvian students

Johana Gonzales-Valdivia[1], Wilter C. Morales-García[2,3]*, Jacksaint Saintila[4], Salomón Huancahuire-Vega[5], Mardel Morales-García[1] and Percy G. Ruiz Mamani[6]

[1]Unidad de Posgrado de Educación, Escuela de Posgrado, Universidad Peruana Unión, Lima, Peru, [2]Unidad de Posgrado de Salud Pública, Escuela de Posgrado, Universidad Peruana Unión, Lima, Peru, [3]Facultad de Teología, Universidad Peruana Unión, Lima, Peru, [4]Escuela de Medicina Humana, Universidad Señor de Sipán, Chiclayo, Peru, [5]Dirección General de Investigación, Universidad Peruana Unión, Lima, Peru, [6]Escuela Profesional de Enfermería, Facultad de Ciencias de la Salud, Universidad Privada San Juan Bautista, Chorrillos, Peru

Risk behaviors in schoolchildren can negatively influence and affect their wellbeing. Therefore, it is necessary to have a validated tool to measure student satisfaction in specific areas of the school. The objective of this study was to adapt to Spanish and evaluate the psychometric properties of the H-Sat Scale in Peruvian students. The participants were 691 high school students between 11 and 18 years old (M = 13.96, SD 1.38) who were administered the H-Sat Scale. The scale was translated into Spanish (H-SatP Scale) using the forward and backward method. The scale presented adequate internal consistency for each of the five factors (ordinal α, CR, ω, and H $>$ 70). Confirmatory analysis confirmed the five-dimensional structure (χ^2 = 620.864, df = 160; CFI = 0.982, TLI = 0.978, RMSEA = 0.065, SRMR = 0.032). This measurement tool could be used for the evaluation of interventions in school and health contexts to assess other aspects of wellbeing necessary for their development in school-age students.

KEYWORDS
high-school, satisfaction, validity, reliability, translation

Introduction

Risky behaviors in schoolchildren can affect physical and mental wellbeing. The school environment affects the health-related behaviors of adolescents both positively and negatively, by establishing behavioral patterns and attitudes, since schools are where they spend the most time. However, little attention is paid to school satisfaction, since it is usually focused on academic performance and little attention is paid to evaluative and affective outcomes (Huebner and McCullough, 2000; Takakura et al., 2010; Wang et al., 2019). The promotion of mental health improvement provides a better life trajectory for the adolescent and a better performance in society (Becker et al., 2019; Fazel and Kohrt, 2019). Satisfaction with school is an important aspect of quality of life, since schoolchildren have the right to feel good about the institutions in which they find themselves. Schools should be environments where children are supported, valued, and enjoyed, since the level of satisfaction affects psychological wellbeing and school engagement and reduces the rate of absenteeism, behavioral problems, and dropout (Verkuyten and Thijs, 2002; Virtanen et al., 2019).

The environment in which schoolchildren develop influences their adaptation and adjustment, because school experiences and self-perception can affect the course of their lives (Baker et al., 2003; Hong et al., 2022). Schools should be evaluated in a broader sense, i.e., not only on academic outcomes but also on positive non-academic outcomes, such as affective issues, student perception of the quality of school environments, and their psychological wellbeing (Huebner and McCullough, 2000; Huebner et al., 2009; Torres-Zapata et al., 2022). Schooling has presented a narrow vision by downplaying the importance of the student's adaptive sense, since the student must be committed to and enjoy what he/she does (Huebner and McCullough, 2000). School satisfaction has been defined as the cognitive-affective assessment of overall satisfaction with school experiences as a whole (Huebner, 1994; Wong and Siu, 2017). Studies indicate that school satisfaction is an important variable because it is associated with learning, behavior, dropout, absenteeism, teacher-student relationship, engagement, and social environments by incurring positive behavioral changes (Okun et al., 1990; Konu and Rimpelä, 2002; Johnson et al., 2022). Likewise, school satisfaction presents subjective or student-perceived indicators that the general evaluation of overall experiences in schoolchildren. It also presents objective observable measures, which refer to the external school domain such as academic results or the positive relationship between students and teachers (Zullig et al., 2011). Therefore, school satisfaction measures are necessary, as they allow for the evaluation of students in specific areas of the school.

Few instruments have been developed that assess satisfaction in a school context such as the Multidimensional Students' Life Satisfaction Scale (MSLSS) (Huebner, 1994), Quality of School Life (QSL) (Epstein and Mcpartland, 2016), and QLS (Karatzias et al., 2001) that assess students during the school period, however, none of the scales incorporates an assessment of emotional experiences at school. The questionnaire for the evaluation of school wellbeing and identification of risk factors (QBS 8-13) (Tobia and Marzocchi, 2015) has also been developed, which evaluates the wellbeing of children and adolescents from 8 to 13 years of age from three perspectives: the student himself, the parents and teachers. The wellbeing model (Konu et al., 2002) has three dimensions: school conditions, social relationships, means for self-fulfillment, and health status. While the Adolescents' Subjective WellBeing in School Scale (ASWBSS) is a measure that incorporates cognitive (school satisfaction) and affective components (positive and negative affect) in an integral way, since it provides six dimensions (achievement, school management, relationships between teachers-students, peer relations, teaching, and academic learning) that refer to specific subdomains of school life (Tian, 2008). Likewise, the Brief Adolescents' Subjective WellBeing in School Scale (BASWBSS) was developed, which follows the design of the ASWBSS, and involves two dimensions of school life (school satisfaction and affection at school) (Tian et al., 2014).

Unlike the other measures, the H-Sat Scale is a brief and multidimensional measure, which presents indicators of satisfaction with life and adds areas related to the choice and usefulness of the school for their professional career. The College Satisfaction Scale (CSS) (Lodi et al., 2017) was first developed in Italian university students, later it was adapted and validated in Italian high school students (H-Sat Scale) by Lodi et al. (2019). This scale evaluates five areas, identified from a review of the literature: appropriateness of choice, quality of school services, relationships with classmates, effectiveness of their study habits, and usefulness for the future career. The psychometric results were adequate and suggest being applied in other contexts and evaluated with other psychological variables.

To date, there is no measure in Spanish that assesses satisfaction with school validated in a school population. Therefore, it is necessary to have measures that show validity and reliability to measure satisfaction with secondary school in Peru or in other Latin American countries. For this reason, this study aimed to adapt and evaluate the psychometric properties of the Spanish version of the H-Sat Scale in Peruvian students.

Methodology

Study design and participants

The study is methodological. The participants were secondary school students from the city of Metropolitan Lima, Peru. The sample size was estimated using the Soper software (Soper, 2020) that considers the anticipated effect size ($\lambda = 0.3$), the statistical power level ($1-\beta = 0.95$), and the desired statistical significance ($\alpha = 0.05$). According to the number of variables observed and latent in the model, resulting in a minimum required sample of 223 participants. A non-probabilistic sample was carried out for the selection of the sample. The inclusion criteria were: (1) to be enrolled in the selected institutions, (2) to have been born in Peru, and (3) to accept informed consent. Students who did not provide informed consent were not considered. The final sample consisted of a total of 691 schoolchildren from five Peruvian educational institutions.

Instruments

La H-Sat Scale (Lodi et al., 2019) was adapted from the Italian College Satisfaction Scale (CSS) (Lodi et al., 2017). It is composed of 20 items that assess student satisfaction in five dimensions through a five-point Likert scale (1 = not at all, 2 = a little, 3 = somewhat, 4 = very, and 5 = completely). The items correspond to 5 dimensions of school satisfaction: appropriateness of choice (CH), quality of school services (SE), relationships with classmates (RE), effectiveness of study habits (TS), and usefulness for a future career (AC). This instrument has good psychometric properties, in terms of validity and internal consistency for the five subscales (CH = α 0.86; SE = α 0.76; RE = α 0.88; TS = α 0.87; AC = α 0.92).

The English to Spanish translation of the H-Sat was carried out using cultural adaptation-based procedures (Beaton et al., 2000).

1. The H-Sat was initially translated independently into Spanish by two bilingual native Spanish speakers. The versions were compared, and an initial version was developed.
2. The Spanish version was again independently translated into English by two Native Americans who speak Spanish but were not familiar with H-Sat.
3. Subsequently, a panel of two educators and two psychologists reviewed the Spanish-translated version and the translated versions and developed the preliminary version of the H-SatP.
4. The preliminary version was administered to a focus group of 15 students from first to fifth grade, to assess comprehension and readability. Apparent comprehension problems were identified; therefore, linguistic changes were made in the final Spanish version (**Table 1**).

Data analysis

Confirmatory factor analysis (CFA), descriptive analysis of the H-SatP items by calculating the mean, standard deviation, skewness, kurtosis, and corrected item test correlation using the packages "*lavan*," "*psych*," and "*Sem Tools*" of R software. For skewness (g1) and kurtosis (g2), values between $\pm$ 1.5 were considered adequate (Pérez and Medrano, 2010), and corrected item-test correlation analysis was considered for item removal in case of r (i-tc) $\leq$ 0.2 or multicollinearity (i-tc) $\leq$ 0.2 (Kline, 2016) and internal consistency were estimated using ordinal α coefficient. The five-factor structure was confirmed from a previous study (Lodi et al., 2019) and the analysis was performed using the weighted least-square method (WLSM) due to the categorical nature of the items, which also allows the detection of structural relationships with slight or moderate asymmetry (Brown, 2015; Li, 2016). The chi-square test (χ^2), confirmatory fit index (CFI $\geq$ 0.95), Tucker–Lewis index (CFI $\geq$ 0.95) (Schumacker and Lomax, 2016), the root-mean-square error of approximation (RMSEA $\leq$ 0.05), and the standardized root-mean-square residuals (SRMR $\leq$ 0.05) were considered for the evaluation of the fit models (Kline, 2016). The evidence of convergent validity was obtained from the Average Variance Extracted (AVE) average variance extracted from the factor, which considers quantities greater than 0.50 as satisfactory (Fornell and Larcker, 1981). In addition, item permanence was determined by means of factor loadings (λ > 0.70) (Dominguez-Lara, 2018).

The internal consistency was analyzed through the ordinal α coefficient (Pascual-Ferrá and Beatty, 2015) composite reliability (CR) (Nunnally and Bernstein, 2010), the coefficient ω (McDonald, 1999), and H (Hancock and Mueller, 2001) expecting high magnitudes (> 0.80) (Raykov and Hancock, 2005; Dominguez-Lara, 2016).

Procedure

The study was approved by the ethics committee of the Universidad Peruana Unión (CE-EPG 000016). Authorization was requested from the directors of three educational institutions to proceed with the collection of information. The evaluation dates were set, and the students' email addresses were accessed, since the students were at home because of the confinement due to the Covid-19 pandemic, their participation was requested via email. institutional. Based on this, a questionnaire was sent through an online Microsoft Forms form in which the researchers explained the research protocol to the parents or guardians, after acceptance by their relatives, the students had access to informed consent. The questionnaire was completed by those students who voluntarily agreed to participate in the study and whose parents approved or signed the informed consent form. The study was conducted from May 30 to June 23, 2021.

Results

Demographic characteristics

The ages of the students ranged between 11 and 18 years (*M* = 13.96, *SD* 1.38), where 50.4% were men and 49.6% were women. On the other hand, 24.7% were in first grade or year of high school, 22.3% in second grade, 19.7% in third grade, 21.6% in fourth grade, and only 11.7% in fifth grade.

Descriptive statistics of H-Sat items

The descriptive statistics are shown in **Table 1**, where item 10 (M = 4.12) has the highest mean, while the lowest mean was found in item 13 (M = 3.63). This same item 13 (SD = 1.22) showed greater dispersion with respect to variability. Skewness (g1) and kurtosis (g2) fluctuated between values below ± 1.5 in all items, indicating a normal multivariate distribution. Likewise, the scale presents item-total correlations between 0.66 and 0.86, higher than the acceptable limit of 0.30, indicating a high homogeneity and the internal consistency for each item was acceptable (>0.95) (see **Table 1**).

Evidence of validity related to internal structure

The CFA was performed by hypothesizing the five-factor model initially proposed by Lodi et al. (2019). The goodness-of-fit indices for the total sample were [χ^2 = 620.864, gl = 160; CFI = 0.982, TLI = 0.978, RMSEA = 0.065 (90% CI: 0.059–0.070), SRMR = 0.032], indicating that the Peruvian version model fits the observed data adequately. In addition, all the λ were greater than 0.70 and the AVE values are adequate (AVE > 0.50), indicating that the latent factors are adequately explained by their observed variables. Furthermore, in terms of reliability, the values obtained were high (ordinal α, CR, ω and H > 70) for all dimensions (see **Table 2**).

Discussion

School satisfaction, according to current positive approaches, is considered an important aspect of adolescents' wellbeing and quality of life (Baker and Maupin, 2009). It is therefore crucial for schools to create appropriate conditions and strategies for students to experience enthusiasm, happiness, and pleasant experiences, which are reflected in a positive attitude toward school and better overall psychological wellbeing (Saminathen et al., 2019). Therefore, it is necessary to have a validated and appropriate instrument to measure and monitor students' school satisfaction levels. The aim of the present research was to translate and adapt the High-School Satisfaction Scale (H-Sat Scale) to Peruvian Spanish (H-SatP) and to evaluate its psychometric properties to measure school satisfaction from a multidimensional perspective.

TABLE 1 Descriptive statistics and reliability.

Item	I am satisfied:	M	SD	g_1	g_2	r.cor	α
CH1	For choosing this school.	4.09	0.94	−0.87	0.19	0.73	0.96
SE2	With classes.	3.67	1.03	−0.33	−0.62	0.74	0.96
RE3	With my classmates' relationships.	3.72	1.08	−0.49	−0.59	0.62	0.96
TS4	With my way of studying.	3.70	1.09	−0.52	−0.49	0.74	0.96
AC5	Because I believe that my courses of study will be useful for my academic and/or professional future.	4.11	1.06	−1.09	0.46	0.80	0.96
CH6	Because I like what I study in this school.	3.95	0.96	−0.74	0.01	0.86	0.96
SE7	With my school's virtual/technological resources.	3.68	1.11	−0.51	−0.62	0.68	0.96
RE8	Because I have classmates with whom I feel comfortable studying.	3.84	1.13	−0.74	−0.31	0.71	0.96
TS9	Of the academic goals I am achieving.	3.88	1.04	−0.67	−0.24	0.83	0.96
AC10	Because I believe that my studies will have a positive effect on my future educational and/or professional career.	4.12	1.00	−1.06	0.56	0.82	0.96
CH11	For starting the school year.	3.91	1.02	−0.79	0.15	0.77	0.96
SE12	With quality of school services (tutoring, sports, among others)	3.80	1.08	−0.67	−0.22	0.80	0.96
RE13	Because I can count on the help of my classmates.	3.63	1.22	−0.52	−0.74	0.66	0.96
TS14	Of my motivation for studying.	3.82	1.07	−0.61	−0.45	0.81	0.96
AC15	Because this school is the basis for my future professional career.	3.98	1.10	−0.9	0.01	0.84	0.96
CH16	Because, after all, this school seems to be tailor-made for me.	3.71	1.09	−0.48	−0.61	0.80	0.96
SE17	With the availability of those who work in the school.	3.86	1.00	−0.61	−0.33	0.77	0.96
RE18	For the good friendships with my classmates.	3.84	1.18	−0.8	−0.33	0.67	0.96
TS19	With my academic performance.	3.73	1.03	−0.47	−0.55	0.69	0.96
AC20	Because attending this school will be helpful in finding future employment.	3.95	1.09	−0.84	−0.1	0.81	0.96

CH, suitability of choice; SE, quality of school services; RE, relations with classmates; TS, effectiveness of study habits; AC, utility for a future career; α, ordinal α .

TABLE 2 Factor loadings and reliability.

	F_1 (λ)	F_2 (λ)	F_3 (λ)	F_4 (λ)	F_5 (λ)
CH1	0.753				
CH6	0.888				
CH11	0.796				
CH16	0.820				
SE2		0.776			
SE7		0.712			
SE12		0.841			
SE17		0.812			
RE3			0.791		
RE8			0.904		
RE13			0.837		
RE18			0.859		
TS4				0.798	
TS9				0.906	
TS14				0.883	
TS19				0.749	
AC5					0.880
AC10					0.903
AC15					0.915
AC20					0.881
AVE	0.67	0.62	0.72	0.70	0.80
Ordinal α	0.89	0.86	0.91	0.90	0.94
CR	0.89	0.87	0.91	0.90	0.94
ω	0.89	0.87	0.91	0.90	0.94
H	0.90	0.87	0.92	0.92	0.94

λ, factor loadings; F1, suitability of choice (CH); F2, quality of school services (SE); F3, relations with classmates (RE); F4, effectiveness of study habits (TS); F5, utility for a future career (AC); composite reliability (CR), ω, omega; H, H-coefficient; AVE, average variance extracted.

Other instruments that measure school satisfaction mostly consider three dimensions: educational quality, teacher support and family support (Rodríguez et al., 2020), or self-determination, including autonomy, competence, and relatedness (Ryan and Deci, 2017). These instruments focus on the student's perception of his or her environment such as teacher and family and the influence they exert on his or her satisfaction. The H-Sat Scale is a reliable, multidimensional, domain-specific measure of students' school satisfaction and has been tested for its relationships with overall life satisfaction. In addition to the personal aspects of satisfaction, the dimensions covered by this scale include satisfaction with their career goals and projects due to the school they attended in terms of programs, knowledge, and skills acquired (Lodi et al., 2019).

Future studies could use the H-Sat Scale in different countries to define whether there are cultural differences with respect to the structural dimensions of school satisfaction and to confirm the psychometric properties of the instrument. The original authors of the H-Sat scale recommended reviewing cultural differences with respect to the structural dimensions of school satisfaction and confirming the psychometric properties of the instrument (Lodi et al., 2019).

In relation to the confirmatory factor analysis, the results were consistent with the original structure. The high magnitudes of the factor loadings were moderate and high and indicate a robust factor structure, with adequate indicators for the construct of satisfaction with high school. Likewise, the values of the Spanish version were close to and higher than the Italian version of the H-SatP. The version produced coefficients of reliability ordinal α, composite reliability (CR), omega (ω), and coefficient H higher than recommended (> 0.70), these results are due to the fact that the coefficients are evaluated through the factor loadings of the latent variables. Therefore, the H-SatP scale is a reliable tool, supporting previous studies. The psychometric results supported the existing version and indicate that the 20 items of the H-SatP are grouped in a five-factor structure: (a) appropriateness of choice, (b) quality of school services, (c) relationships with classmates, (d) effectiveness of study habits, and (e) usefulness for a future career (Lodi et al., 2019).

The first factor, appropriateness of choice, relates to the student's perception of satisfaction based on the evaluation of the service received, whether or not their expectations were met with the services offered (Petruzzellis and Romanazzi, 2010). In other words, students assume the role of the customer who decides to choose the product and whose satisfaction will be reflected in the fulfillment of their expectations. The managers of educational institutions should be aware of the factors that improve the perception of student satisfaction, so that they can provide better services and improve existing ones. Regarding the second factor, different studies show the influence of school service quality on school satisfaction (Patrón-Cortés, 2021). Service quality in turn encompasses aspects such as reliability, security, empathy, responsiveness, and tangibles (Zeithaml et al., 2010). Of all these, reliability is critical in the education sector, as it means receiving the service as promised (Usman, 2010). Therefore, to ensure school satisfaction it is important to make provisions for immediate corrective actions when faults occur, to regain students' confidence and ensure their satisfaction.

Feeling peer support and perceiving a positive school environment are aspects considered in the third factor. A good relationship with classmates is a factor that promotes school wellbeing (Ruvalcaba-Romero et al., 2018). A starting point for the acquisition of behavioral patterns and the establishment of emotional state during adolescence is the quality of relationships experienced in school settings (Plazas et al., 2010). Achieving satisfactory relationships with peers is associated with greater emotional wellbeing and satisfaction with life, which is why social interactions represent one of the most significant school relationships (Uslu and Gizir, 2017). On the other hand, the relationship between academic satisfaction and the effectiveness of study habits can be confirmed by previous studies (Alzahrani et al., 2018). The positive relationship between study habits and academic achievement would reflect even better the satisfaction experienced by schoolchildren. The last factor indicates the positive relationship between usefulness for a future career and school satisfaction; this relationship was confirmed in studies that also related accumulated grades, level of achievement, and satisfaction (El-Hilali et al., 2015). The perceived usefulness of school studies for future university studies should be an aspect to be addressed in orientations and workshops by educational institutions, so that students can confirm the importance of their school studies for their professional future and in turn guarantee academic satisfaction.

The promotion of wellness in school practice should have a growing commitment to service delivery, students should benefit from programs that allow the improvement of their skills to develop competencies, and that contribute to a better social climate in schools. In that sense, the H-SatP scale could be used in environments with complex and multifaceted school entities to understand the potentially reciprocal and hierarchical nature embedded in school environments. The advantage of this scale is its multidimensionality despite having a reduced number of items. In addition, it highlights the fundamental role of the school's influence in achieving student wellbeing. Therefore, it is also a useful instrument for schools as a self-evaluation tool in the quality of service provided.

Despite being the first attempt to evaluate the psychometric properties of H-SatP, it had limitations such as the sample was taken by convenience. Test-retest reliability was not considered, nor was factorial invariance, and it was a cross-sectional study. Despite the limitations, the results are important for the educational part, given that from a theoretical point of view the study provides a multidimensional tool. In addition, it could be used for the evaluation of interventions in school and health contexts to assess other aspects of wellbeing necessary for their development in school-age students.

Conclusion

The Spanish version of the H-SatP has adequate psychometric evidence. This Spanish version allows us to consider psycholinguistic variations in order to have a culturally appropriate interpretation.

Data availability statement

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

Ethics statement

The study was approved by the Ethics Committee of the Universidad Peruana Unión (CE-EPG 000016). Written informed consent to participate in this study was provided by the participants' or their legal guardian/next of kin.

Author contributions

WM-G, PR, and JG-V were in charge of the conceptualization. JS, MM-G, and SH-V participated in the development of the methodology. WM-G and PR participated in the development of the software. WM-G, MM-G, and PR contributed to the formal analysis and research. PR, WM-G, and SH-V carried out writing, review, and editing of the first draft. MM-G, JS, and JG-V carried out visualization and supervision. All authors approved the final version of the 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.

References

Alzahrani, S. S., Soo Park, Y., and Tekian, A. (2018). Study habits and academic achievement among medical students: A comparison between male and female subjects. *Med. Teach.* 40, S1–S9. doi: 10.1080/0142159X.2018.1464650

Baker, J. A., Dilly, L. J., Aupperlee, J. L., and Patil, S. A. (2003). The developmental context of school satisfaction: Schools as psychologically healthy environments. *Sch. Psychol. Q.* 18, 206–221. doi: 10.1521/SCPQ.18.2.206.21861

Baker, J. A., and Maupin, A. N. (2009). "School satisfaction and children's positive school adjustment," in *Handbook of Positive Psychology in Schools*, eds M. J. Furlong, R. Gilman and E. S. Huebner (Abingdon: Taylor and Francis), 189–196.

Beaton, D. E., Bombardier, C., Guillemin, F., and Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25, 3186–3191. doi: 10.1097/00007632-200012150-00014

Becker, K. D., Park, A. L., Boustani, M. M., and Chorpita, B. F. (2019). A pilot study to examine the feasibility and acceptability of a coordinated intervention design to address treatment engagement challenges in school mental health services. *J. Sch. Psychol.* 76, 78–88. doi: 10.1016/J.JSP.2019.07.013

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*, 2nd Edn. New York, NY: Guilford Press.

Dominguez-Lara, S. (2018). Proposal for cut-offs for factor loadings: A construct reliability perspective. *Enfermería Clín.* 28, 401–402. doi: 10.1016/j.enfcle.2018.06.006

Dominguez-Lara, S. A. (2016). Evaluación de la confiabilidad del constructo mediante el Coeficiente H: breve revisión conceptual y aplicaciones. *Psychologia* 10, 87–94.

El-Hilali, N., Al-Jaber, S., and Hussein, L. (2015). Students' satisfaction and achievement and absorption capacity in higher education. *Proc. Soc. Behav. Sci.* 177, 420–427. doi: 10.1016/J.SBSPRO.2015.02.384

Epstein, J. L., and Mcpartland, J. M. (2016). The concept and measurement of the quality of school life. *Am. Educ. Res. J.* 13, 15–30. doi: 10.3102/00028312013001015

Fazel, M., and Kohrt, B. A. (2019). Prevention versus intervention in school mental health. *Lancet Psychiatry* 6, 969–971. doi: 10.1016/S2215-0366(19)30440-7

Fornell, C., and Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *J. Market. Res.* 18, 382–388. doi: 10.1177/002224378101800313

Hancock, G. R., and Mueller, R. O. (2001). "Rethinking construct reliability within latent variable systems," in *Structural Equation Modeling: Present and Future—A Festschrift in Honor of Karl Joreskog*, eds R. Cudeck, S. D. Toit, and D. Soerbom (Lincolnwood, IL: Scientific Software International), 195–216.

Hong, X., Lu, Y., and Zhu, W. (2022). *Children's School Feelings and Adaptation During the Transition From Kindergarten to Primary School in China*. London: Routledge, doi: 10.1080/10409289.2022.2090774

Huebner, E. S. (1994). Preliminary Development and Validation of a Multidimensional Life Satisfaction Scale for Children. *Psychological Assessment* 6, 149–158. doi: 10.1037/1040-3590.6.2.149

Huebner, E. S., Gilman, R., Reschly, A. L., and Hall, R. (2009). "Positive schools," in *Oxford handbook of positive psychology*, eds C. R. Snyder, and S. J. Lopez (Oxford: Oxford University Press), 561–569. doi: 10.1093/oxfordhb/9780195187243.013.0053

Huebner, E. S., and McCullough, G. (2000). Correlates of school satisfaction among adolescents. *J. Educ. Res.* 93, 331–335. doi: 10.1080/00220670009598725

Johnson, B., Reschly, A. L., Martin, R. P., Pinzone, C., and Appleton, J. J. (2022). *Student Engagement in Elementary School: Profiles and Associated Characteristics*. London: Routledge, doi: 10.1080/10824669.2022.2112196

Karatzias, A., Power, K. G., and Swanson, V. (2001). Quality of school life: Development and preliminary standardisation of an instrument based on performance indicators in Scottish secondary schools. *Sch. Effect. Sch. Improv.* 12, 265–284. doi: 10.1076/SESI.12.3.265.3449

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Press.

Konu, A., Alanen, E., Lintonen, T., and Rimpelä, M. (2002). Factor structure of the school well-being model. *Health Educ. Res.* 17, 732–742. doi: 10.1093/HER/17.6.732

Konu, A., and Rimpelä, M. (2002). Well-being in schools: a conceptual model. *Health Promot. Int.* 17, 79–87. doi: 10.1093/HEAPRO/17.1.79

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949. doi: 10.3758/S13428-015-0619-7/TABLES/6

Lodi, E., Boerchi, D., Magnano, P., and Patrizi, P. (2017). College Satisfaction Scale (CSS): Evaluation of contextual satisfaction in relation to college student life satisfaction and academic performance. *Appl. Psychol. Bull.*279, 51–64.

Lodi, E., Boerchi, D., Magnano, P., and Patrizi, P. (2019). High-school satisfaction scale (h-sat scale): Evaluation of contextual satisfaction in relation to high-school students' life satisfaction. *Behav. Sci.* 9:125. doi: 10.3390/bs9120125

McDonald, R. P. (1999). *Test Theory: A United Treatment*. Mahwah, NJ: Lawrence Erlbaum.

Nunnally, J. C., and Bernstein, I. H. (2010). *Psychometric Theory*, 3rd Edn. New York, NY: McGraw-Hill Education.

Okun, M. A., Braver, M. W., and Weir, R. M. (1990). Grade level differences in school satisfaction. *Soc. Indic. Res.* 22, 419–427. doi: 10.1007/BF00303835

Pascual-Ferrá, P., and Beatty, M. J. (2015). Correcting internal consistency estimates inflated by correlated item errors. *Commun. Res. Rep.* 32, 347–352. doi: 10.1080/08824096.2015.1089858

Patrón-Cortés, R. M. (2021). Calidad y satisfacción de los servicios escolares de una institución de educación superior mexicana: la visión de los estudiantes. RIDE. *Rev. Iberoam. Para Invest. Desarrollo Educ.* 11:189. doi: 10.23913/RIDE.V11I22.870

Pérez, E. R., and Medrano, L. (2010). Análisis factorial exploratorio : Bases conceptuales y metodológicas. *Rev. Argentina Ciencias Comportamiento* 2, 58–66.

Petruzzellis, L., and Romanazzi, S. (2010). Educational value: How students choose university: Evidence from an Italian university. *Int. J. Educ. Manag.* 24, 139–158. doi: 10.1108/09513541011020954/FULL/XML

Plazas, E. A., Cotes, M. L. M., Santiago, A., Sarmiento, H. H., López, S. E. A., and Patiño, C. D. (2010). Relaciones entre iguales, conducta prosocial y género desde la educación primaria hasta la universitaria en Colombia. *Univ. Psychol.* 9, 357–370. doi: 10.11144/JAVERIANA.UPSY9-2.RICP

Raykov, T., and Hancock, G. R. (2005). Examining change in maximal reliability for multiple-component measuring instruments. *Br. J. Math. Stat. Psychol.* 58, 65–82. doi: 10.1348/000711005X38753

Rodríguez, C., Padilla, G., and Gallegos, M. (2020). Calidad educativa, apoyo docente y familiar percibido: la tridimensionalidad de la satisfacción escolar en niños y adolescentes. *Cuadernos Invest. Educ.* 11, 157–173. doi: 10.18861/CIED.2020.11.2.2995

Ruvalcaba-Romero, N. A., Orozco-Solisa, M. G., Gallegos-Guajardo, J., and Nava-Fuerte, J. M. (2018). Relaciones escolares, comunicación con padres y prosocialidad como predictores de emociones positivas. *Liberabit* 24, 183–193. doi: 10.24265/LIBERABIT.2018.V24N2.02

Ryan, R. M., and Deci, E. L. (2017). *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York, NY: Guilford Press, doi: 10.1521/978.14625/28806

Saminathen, M. G., Låftman, S. B., and Modin, B. (2019). School choice at a cost? Academic achievement, school satisfaction and psychological complaints among students in disadvantaged areas of Stockholm. *Int. J. Environ. Res. Public Health* 16:1912. doi: 10.3390/IJERPH16111912

Schumacker, R. E., and Lomax, R. G. (2016). *A Beginner's Guide to Structural Equation Modeling*, 4th Edn. New York, NY: Taylor & Francis.

Soper, D. (2020). *A-priori Sample Size Calculator for Structural Equation Models*. Available online at: https://www.danielsoper.com/statcalc/default.aspx (accessed May 2021).

Takakura, M., Wake, N., and Kobayashi, M. (2010). The contextual effect of school satisfaction on health-risk behaviors in Japanese high school students. *J. Sch. Health* 80, 544–551. doi: 10.1111/J.1746-1561.2010.00540.X

Tian, L. (2008). Developing scale for school well-being in adolescents. *Psychol. Dev. Educ.* 24, 100–106.

Tian, L., Wang, D., and Huebner, E. S. (2014). Development and validation of the brief adolescents' subjective well-being in school scale (BASWBSS). *Soc. Indic. Res.* 120, 615–634. doi: 10.1007/S11205-014-0603-0

Tobia, V., and Marzocchi, G. M. (2015). *QBS 8-13 : Questionari per la Valutazione del Benessere Scolastico e Identificazione dei Fattori di Rischio*. Portland, OR: Erickson.

Torres-Zapata, ÁE., Pérez-Jaimes, A. K., Brito-Cruz, T., del, J., and Estrada-Reyes, C. U. (2022). Rendimiento y clima escolar en la unidad de aprendizaje de bioquímica. *Inform. Tecnol.* 33, 225–234. doi: 10.4067/S0718-07642022000200225

Uslu, F., and Gizir, S. (2017). School belonging of adolescents: The role of teacher-student relationships, peer relationships and family involvement. *Educ. Sci. Theor. Pract.* 17, 63–82. doi: 10.12738/ESTP.2017.1.0104

Usman, A. (2010). The impact of service quality on students' satisfaction in higher education institutes of Punjab. *J. Manag. Res.* 2:418. doi: 10.5296/JMR.V2I2.418

Verkuyten, M., and Thijs, J. (2002). School satisfaction of elementary school children: The role of performance, peer relations, ethnicity and gender. *Soc. Indic. Res.* 59, 203–228. doi: 10.1023/A:1016279602893

Virtanen, T. E., Vasalampi, K., Torppa, M., Lerkkanen, M.-K., and Nurmi, J.-E. (2019). Changes in students' psychological well-being during transition from primary school to lower secondary school: A person-centered approach. *Learn. Individ. Differ.* 69, 138–149. doi: 10.1016/J.LINDIF.2018.12.001

Wang, Y., Tian, L., and Scott Huebner, E. (2019). Basic psychological needs satisfaction at school, behavioral school engagement, and academic achievement: Longitudinal reciprocal relations among elementary school students. *Contemp. Educ. Psychol.* 56, 130–139. doi: 10.1016/J.CEDPSYCH.2019.01.003

Wong, T. K. Y., and Siu, A. F. Y. (2017). Relationships between school climate dimensions and adolescents' school life satisfaction, academic satisfaction and perceived popularity within a Chinese context. *Sch. Mental Health* 9, 237–248. doi: 10.1007/S12310-017-9209-4

Zeithaml, V. A., Bitner, M. J., and Gremler, D. D. (2010). *Services Marketing Strategy. Wiley International Encyclopedia of Marketing*. Hoboken, NJ: John Wiley & Sons, Ltd, doi: 10.1002/9781444316568.WIEM01055

Zullig, K. J., Huebner, E. S., and Patton, J. M. (2011). Relationships among school climate domains and school satisfaction. *Psychol. Sch.* 48, 133–145. doi: 10.1002/PITS.20532

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 21 October 2022
DOI 10.3389/feduc.2022.893829

OPEN ACCESS

EDITED BY
Antonio P. Gutierrez de Blume,
Georgia Southern University,
United States

REVIEWED BY
Yuri Sato,
The University of Tokyo, Japan
Shogo Kajimura,
Tamagawa University, Japan

*CORRESPONDENCE
Hiroaki Ayabe
ayabe@academion.com

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

RECEIVED 10 March 2022
ACCEPTED 23 September 2022
PUBLISHED 21 October 2022

CITATION
Ayabe H, Manalo E, Hanaki N, Fujita H and Nomura M (2022) Brain imaging provides insights about the interaction between instruction and diagram use for mathematical word problem solving.
Front. Educ. 7:893829.
doi: 10.3389/feduc.2022.893829

Brain imaging provides insights about the interaction between instruction and diagram use for mathematical word problem solving

Hiroaki Ayabe[1,2]*, Emmanuel Manalo[1], Noriko Hanaki[3], Hiroyo Fujita[1] and Michio Nomura[1]

[1]Division of Cognitive Psychology in Education, Graduate School of Education, Kyoto University, Kyoto, Japan, [2]Division of Cerebral Integration, Department of System Neuroscience, National Institute for Physiological Sciences, Okazaki, Japan, [3]Department of Psychology and Education, Faculty of Liberal Arts, The Open University of Japan, Chiba, Japan

Diagram use is generally considered an effective strategy in solving mathematical word problems, and many teachers demonstrate the use of this strategy when they are teaching. Despite such demonstrations, however, the majority of students evidence poor, ineffective use of diagrams in problem solving. This may be due to a lack of task-appropriate instruction, but to date, there is inadequate evidence to support this idea. Therefore, the present study aimed to better understand the development of diagram use competence following the provision of task-appropriate instruction, focusing on both behavioral and neurophysiological evidence (i.e., brain activity, using functional near-infrared spectroscopy or fNIRS). Sixteen participants (mean age 15.7 years) were asked to solve mathematical word problems for which the use of tables (which is one kind of diagram) was deemed effective. Data collection progressed in three phases: (1) Pre-test without the demand for diagram use, (2) Pre-test with demand to use a table, and (3) Post-test (after participants received instruction on table use for problem solving). Although table use increased in Phase 2, it was only in Phase 3 that such use led to increases in correct answers. In Phase 3, fNIRS measurements also indicated an increase in blood flow to the frontal area (DLPFC and VLPFC) of the brain usually associated with working memory activity. These results demonstrate important neurophysiological changes resulting from task-appropriate instruction that promotes effective strategy use and improves learning performance.

KEYWORDS

self-constructed diagrams, mathematics education, problem solving, tables and arrays, brain imaging, functional near-infrared spectroscopy (fNIRS), learning strategy use, working memory

Introduction

"A factory manufactures automobile parts, packing 150 parts in each box, and carries as many boxes as possible to an assembly plant every day. If the number of parts left is more than 75, the factory will produce 180 parts the next day. Otherwise, it will produce 220 parts. The number of the parts left on day 1 was 130. How many parts will be left on day 365?"

Even elementary school students can probably reach the correct answer to this word problem by repeated calculations. However, it would take a tremendous amount of effort and time to do that. Mathematical word problems (often abbreviated as just word problems or WPs) are taught at all levels of education to develop the ability to apply mathematical knowledge to the real world. The word problems usually represent the relevant information in a short narrative rather than mathematical notation. For solving word problems efficiently, diagrams play an important role. However, teachers tend to face difficulty in getting students to master strategies for using diagrams. This study aimed to shed light on developing students' ability to use diagrams that are appropriate for solving corresponding types of word problems. To deepen our understanding of the mechanisms involved in the development of such ability, we included a neurophysiological approach using functional near-infrared spectroscopy (fNIRS) to the behavioral approach mainly utilized in psychology and educational science.

Mathematical word problem solving and diagram use

It is necessary for us to use mathematics in various facets of everyday life, from making proportional adjustments to amounts of ingredients when cooking to planning travel activities. For this reason, developing skills in solving mathematical word problems is important because such problems contextualize the need to apply mathematical knowledge and skills, rendering them useful practice for real-life situations. Therefore, it is important that instruction provided in school is effective in developing such skills. The school curriculum in most countries normally includes mathematical word problem solving. However, previous research studies have reported that not only students, but people in general are not good at solving such problems (Riley et al., 1983; Lewis and Mayer, 1987; Hegarty et al., 1992; Daroczy et al., 2015).

Mathematical word problems have been portrayed as being one of the most difficult and complex types of problems that students encounter during their mathematics education (Daroczy et al., 2015). The difficulties arise from (i) the linguistic complexity of the text describing the problems, (ii) the numerical complexity of the mathematical problems themselves, and (iii) the complexity of the linguistic and numerical relationships portrayed in those problems. The knowledge and skills necessary for managing the third of these sources of difficulties are not usually explicitly taught in any school subject as they lie in the interaction between knowledge domains (in this case, between language and mathematics). However, in modern societies like those that most of us inhabit in the current century, competence in dealing with interactions between knowledge domains is crucial because real-world problems will present similar complexities. This is why training in reality-emulating problem solving—like mathematical word problem solving—is so important (Education 2030; OECD, 2018).

Diagrams are considered effective to use when solving mathematical word problems (Larkin and Simon, 1987; Novick and Hurley, 2001; García et al., 2006; Ainsworth et al., 2011; Boonen et al., 2016). A meta-analysis that included 487 studies reported that diagrams can be considered one of the most effective heuristics to use for such problem solving (Hembree, 1992). Mathematics teachers in general are knowledgeable about the effectiveness of diagram use, and they often demonstrate how to use diagrams when teaching. Diagrams can help their students understand the linguistic and numerical relationships present in mathematical word problems (Van Garderen, 2007; Daroczy et al., 2015). However, despite such demonstrations of diagram use, students tend not to use diagrams when solving mathematical word problems and, even when they do use diagrams, they often fail to derive the correct answers (Uesaka and Manalo, 2006; Uesaka et al., 2010). There are at least two problems that exist.

Problems in diagram use

The first problem is that students tend not to spontaneously use diagrams for solving mathematical word problems (e.g., Uesaka et al., 2007; Uesaka and Manalo, 2012). Apart from simply not knowing how to use diagrams, one important reason that has been identified regarding this problem concerns cognitive load. Understanding the text of the mathematical word problem and the terms of the problem that it conveys is already cognitively demanding (Sweller et al., 1998; Sweller, 2010; Paas et al., 2011; Schmeck et al., 2015). However, the procedure for translating the pertinent parts of the mathematical word problem into a diagram is an additional high load that students need to bear (Uesaka and Manalo, 2012). The resulting high cognitive load involved in using a diagram during the problem solving process can make such use prohibitive for students (Uesaka and Manalo, 2012; Ayabe and Manalo, 2018). Thus, high cognitive load may lead to production deficiencies in utilizing diagrams as a strategy (cf. Bjorklund et al., 1997, where strategy use in general is concerned).

The second problem is that students cannot always solve the mathematical word problem successfully despite using what could be considered the appropriate kinds of diagram (Larkin and Simon, 1987; Mayer and Gallini, 1990). Instruction on the use of diagrams have been indicated in previous research as effective for encouraging students to create the necessary kinds of diagrams for the problems given, but the construction of those diagrams does not always lead to the production of correct answers (Uesaka and Manalo, 2006; Uesaka et al., 2010). The results indicate that only creating the appropriate kind of diagram is not enough for students to solve the mathematical word problem successfully: they need to also be able to include the necessary problem-specific numerical and other details in those diagrams in order to work out the answers, as well as to be able to draw the necessary problem-specific inferences from those diagrams (e.g., Cartesian graphs enable working out the relative changes in two quantities usually across time, and tables are appropriate for discovering the rule that governs predictable change in quantities). Therefore, inadequate problem-appropriate details included in instruction on diagram use may lead to such a utilization deficiency in using diagrams as a strategy (cf. Bjorklund et al., 1997; Reuter et al., 2015).

To consider why the utilization deficiency occurs, we need to focus on whether the diagram constructed contains the numerical and other details necessary to meet the cognitive demands of the problem. Any problem can have multiple alternative forms of external representation, drawing attention to different features and providing "representational guidance" (Suthers, 2003; Schnotz and Kürschner, 2008). For example, most students can create a visual-schematic representation of a problem situation, such as illustrations or schematic diagrams. The use of such diagrams has been demonstrated in previous studies to lead to the generation of correct solutions in problem solving (Hegarty and Kozhevnikov, 1999; Boonen et al., 2014). Their purpose, however, is simply to visually represent the problem situation in a more-or-less concrete form (e.g., where objects are located in relation to each other), so that the problem solver can understand how it may be possible to solve it. We define the knowledge necessary to use such external representations as "general knowledge" (about diagrammatic representations). Such knowledge can be used to approach solutions to a wide range of word problems. However, general knowledge about the use of diagrams (like in this example) is often inadequate for solving more complex problems. Those types of problems usually require more abstract forms of diagrams to solve: concrete representation of the problem situation is usually not enough, and the diagram itself needs to directly facilitate the working out of the solution. For such problems, specific diagram knowledge is necessary. Specific diagram knowledge enables matching of the requirements of the problem to the "problem-appropriate diagram" that needs to be used. The important point to note here is that, when solving more complex and/or more abstract problems, usually not any diagrammatic representation will suffice: a specific kind of diagram needs to be used correctly to enable the appropriate operations for solving the problem to be executed (cf. Zhang and Norman, 1994; Zhang, 1997; Duval, 2006).

The problem is that in many classroom settings, when teachers simply demonstrate the use of diagrams such as tables to solve more complex mathematical word problems, they do not provide sufficient instruction to cultivate the problem-appropriate knowledge necessary for the correct construction and use of such diagrams (JSME, 2009). Thus, many students only apprehend the general aspects of the abstract diagram: for example, they understand that tables can be used to solve complicated problems, but they do not know the exact details of how to construct and use tables correctly for such purposes.

Purpose and rationale of the present study

The present study had a twofold purpose. First, it examined whether the provision of problem-appropriate instruction might help develop diagram use competence that would result in overcoming the two problems noted in the previous subsection. In other words, that both spontaneity in appropriate diagram use and generation of correct answers would be promoted. Second, it sought not only behavioral evidence, but also neurophysiological evidence that the cognitive mechanisms that drive the *ability to simply construct a diagram* and the *ability to correctly construct and use a diagram according to task specific requirements* are distinct and identifiable. This study used arithmetic or numeric mathematical word problems, which are only a small part of the types of problems available.

Previous studies have reported that declarative knowledge ("what to use"), procedural knowledge ("how to use"), and conditional knowledge ("when to use") are all necessary to successfully use learning strategies (Brown et al., 1981; Paris et al., 1983; Garner, 1990). When teachers explain how to solve a particular mathematical word problem through the use of a diagram, they may generally believe that demonstration of such use is sufficient. However, it may only largely promote the acquisition of declarative knowledge (i.e., what to use = diagram strategy) and, although both procedural knowledge and conditional knowledge would be implicit in such a demonstration, they may not be adequately salient or detailed enough to facilitate acquisition. Instruction therefore needs to include sufficient details of the steps to take in correctly using the strategy in specific types of problems, and explicit explanations of the conditions for when to use the strategy (including specific steps that may only apply when certain conditions are present in the problem given). On top of such instruction, practice would be indispensable: it would help consolidate procedural and conditional knowledge necessary for correctly using the strategy (diagrams in this case) for the specific types of problems being considered.

If instruction is provided not in a broad-brush manner but specifically, dealing with details of the types of word problems that students should be able to solve and the ways that appropriate kinds of diagrams can be constructed and used to solve those problems, the understanding of the problem-appropriate diagram can effectively be conveyed to students. Combined with adequate practice with feedback, students should then be able to not only overcome their lack of spontaneity in using diagrams, but also the problem of failing to generate the correct answers despite diagram use. They would not only know that the problem-appropriate diagram would help in solving the problem—and hence spontaneously use it, but also how to construct and use that diagram to help obtain the answer required.

The solution proposed above needs to be tested and verified, and doing so requires not only measurements of problem solving behavior and performance, but also of the cognitive process involved in changes that are proposed to occur. It is often difficult to ascertain when students have acquired problem-appropriate diagram knowledge as opposed to general knowledge about diagram use based on behavioral and performance measures alone as those do not objectively portray the changes that occur in the minds of students. Recently, however, physiological studies have clarified the various localization of brain functions. Therefore, we believe it useful to examine whether there might be neurophysiological correlates of the changes in behavior/performance that we were hypothesizing would occur as a consequence of providing problem-appropriate diagram use instruction. Evidence of such correlates would confirm that such instruction leads to enabling at least some of the cognitive processes that are necessary for students to develop.

As mentioned earlier, the cognitive processing involved in solving mathematical word problems is complex. However, focusing on the prefrontal cortex will ascertain whether diagram use works successfully. Solving mathematical word problem includes at least numerical and linguistic processing. Numeric processing involves the frontal-parietal network (Dehaene et al., 1999; De Smedt et al., 2011), and linguistic processing involves the frontal, temporal, and parietal network (Sugiura et al., 2011). Researchers have revealed more detailed brain connections in recent years (Peters and De Smedt, 2018; Vogel and De Smedt, 2021), but significantly, activation in the prefrontal cortex underlies cognitive processing in common (Soltanlou et al., 2018). One of the mechanisms that mediate the prefrontal cortex's hemodynamics and cognitive mechanisms is the working memory system, a central theory in cognitive science.

Working memory comprises a cognitive system that integrates (in what has been called the "episodic buffer") and processes (in its "executive function") visual information (in what has been called the "visuospatial sketch pad") and verbal information (in what has been called the "phonological loop") while temporarily retaining those kinds of information (Baddeley, 1992, 2000). Solving a mathematical word problem would require the use of working memory related to schema construction and application (Sweller et al., 1998).

The central executive system of working memory and its two sub-systems of the phonological loop and the visuospatial sketch pad have been found to correspond to frontal lobe functions in neuroscience (i.e., the prefrontal cortex of the left and right hemispheres of the brain) (D'Esposito et al., 1995; Smith et al., 1996; Kane and Engle, 2002). The DLPFC (dorsolateral prefrontal cortex) processes endogenous, complex, and higher-order information based on attention and concentration (Corbetta and Shulman, 2002; Petrides, 2005; O'Reilly, 2010), while the VLPFC (ventrolateral prefrontal cortex) processes exogenous stimuli as "simple working memory" content and the functions of selection and comparison as "first-order executive functions." Numerical (linguistic) processing activates the left DLPFC, and retention and updating of that information—presumably in the phonological loop—activate the left VLPFC (Paulesu et al., 1993; Boisgueheneuc et al., 2006). Spatial (non-linguistic) cognitive processing—which is supposed to be undertaken through the visuospatial sketch pad—activates the right VLPFC and DLPFC (Smith et al., 1996).

Cognitive neuroscientific studies on diagram use also reported that the prefrontal cortex plays an important role. An fMRI study using a simple syllogism task showed lower brain activity in the left PFC (frontal lobe) when presented with an Euler diagram than with text alone. The results suggest that problem-appropriate diagrams assist the visual channels of working memory and facilitate logical reasoning (Sato et al., 2015). On the other hand, an fNIRS study using the Cleveland task (Cleveland and McGill, 1984) comparing pie charts and bar charts showed lower activity in the prefrontal cortex when using diagrams less appropriate to the problem (Peck et al., 2013). These two findings seem contradictory at first glance. However, they are consistent with working memory system theory (cognitive load theory; CLT; Sweller, 2010; Sweller et al., 1998) because we can consider that the former was due to a decrease in extrinsic cognitive load and the latter to a failure to allocate germane cognitive load. Importantly, these results provide neuroscientific evidence for an association between diagram use and brain activity. Therefore, evaluating behavioral outcomes based on these findings should help identify more detailed cognitive processes.

We selected a near-infrared optical functional brain imaging device (fNIRS) because it can be considered suitable for examining cognitive processes in school children and students when they are solving mathematical word problems with the use of a diagram. Many studies examining mathematical cognitive processes use fMRI, EEG, and fNIRS because they are non-invasive. fMRI has high spatial resolution and has the advantage of being able to measure the whole brain. However, it is sensitive to motion artifacts, which raises the problem of ecological plausibility if attempting to examine cognitive processes that involve movement. MR scanners produce a loud noise and force cramped spaces, putting school children and students in a harsh situation (Skau et al., 2022). EEG, on the other hand, can capture the instantaneous potential of neuronal firing with its high temporal resolution. However, EEG also is sensitive to motion artifacts. EEG analysis usually attempts to remove artifacts widely ranging from eye movements to faulty electrodes (Reis et al., 2014; Jiang et al., 2019). Therefore, unexpected and sudden movements cause critical problems. The availability of mobile EEG for practical use has recently improved (Biondi et al., 2022). However, most mobile devices are purpose-designed with low density (only some electrodes, including Fp1/Fp2). Many of those employ dry electrodes, which are more sensitive to motion artifacts. Accordingly, future validation and standardization must improve signal quality, validity, and reliability (Lau-Zhu et al., 2019). For those reasons, using an EEG device, whether conventional or mobile, is challenging for this study focusing on complex cognitive activity when school children and students solve for a number of minutes in an erratic, unpredictable manner (Xu and Zhong, 2018). Finally, fNIRS signals can show long-lasting changes to examine what happens during mathematical word problem solving (Sakai, 2022). Compared to fMRI, fNIRS has a lower spatial resolution and cannot detect blood flow activity in deeper brain regions (over 1–2 cm). Also, compared to EEG, fNIRS has (a higher spatial resolution but) a lower time resolution. Despite these functional limitations, the ability of fNIRS is sufficient for this study to focus on cerebral blood flow in the prefrontal cortex during students' attempts at solving problems using diagrams. fNIRS is more robust to the motion artifacts than the other methods, enabling the assessment of the neural basis of cognitive processes with movement (Soltanlou et al., 2018). fNIRS is a friendly device even for children and young students to use in a realistic environment with almost no noise (Pinti et al., 2020). fNIRS has the advantage of producing data that are highly correlated with data gathered using fMRI (Cui et al., 2011) and it is both inexpensive and portable.

Main questions and hypotheses

There are a number of questions we wanted to address in this research. The first question was whether students would be able to construct what could be considered problem-appropriate diagrams for a particular type of problem given to them, and whether despite such diagram construction their problem solving performance (i.e., ability to produce the correct answers) would not improve. To answer this question, we use mathematical word problems like the one we introduced at the beginning of this article. We designed them so that the use of a diagram would be necessary. In such problems, students cannot easily imagine how to solve them and use routine algorithmic computation procedures. The use of such problems therefore can help in determining students' diagram knowledge (Reuter et al., 2015). In this case, the kind of diagrams appropriate for the mathematical word problems we designed and administered to the students was tables. We decided on this design because all students in Japan are supposed to learn how to use tables for solving mathematical word problems (Ayabe et al., 2021a). To examine the level of table knowledge students have, the prompt "Use TABLE to solve the word problem" can be helpful. When the prompt is given, the students can create the table because they should know what it is. However, if they do so and they cannot use the table to solve the problem, it would suggest that they possessed adequate general knowledge about tables (i.e., enough to know what they are and how to construct them when asked to do so), but not enough to construct them with the details necessary to meet the requirements of the problem and/or to execute the operations that tables would allow in order to produce the correct answer (Mayer and Gallini, 1990; Reuter et al., 2015).

Our second question was whether students' abilities to produce correct answers would improve after they receive problem-appropriate instruction on the use of tables for solving those problems. If they prove able to do this, it would suggest that such instruction equipped them with problem-appropriate knowledge about table use—in this case, specific to solving the types of problem they were administered.

A third question we had was whether, apart from evidence from problem solving performance, it would be possible to verify such differences in general and problem-appropriate knowledge about the use of a diagram (tables) through neurophysiological evidence (i.e., differences in brain activity). The possession of the relevant general knowledge would likely be adequate to enable the construction of tables. However, with the possession of only such knowledge, students are unlikely to be able to utilize the table they construct to correctly solve the problem they have been given. In such a situation, we predicted the blood flow in the prefrontal cortex (VLPFC and DLPFC) would decrease.

We can verify the prediction by manipulating the prompt described before. If we do not give the prompt, the students are free to answer in any way they choose. They can perform calculations using numbers and conditions in the problem on their own. In addition, they probably would not construct tables (Uesaka et al., 2007; Elia et al., 2009), they can also allocate the germane cognitive load to their surplus cognitive resources in a way they believe (CLT; Sweller et al., 1998; Sweller, 2010). It means the executive function of working memory works actively. Therefore, cerebral blood flow in the prefrontal cortex should be large.

On the other hand, the prompt "Use TABLE to solve the word problem" would restrict free solvers' behavior and cognition. They would struggle if they do not have table knowledge for solving problems. Even if they constructed a table, it would likely remain incomplete because they know the table only in language and general appearance. Most of them should not be able to visualize it completely (Mayer and Gallini, 1990; Reuter et al., 2015). In this case, allocation of germane cognitive load would likely fail and result in the stagnation of the executive function of working memory. Therefore, cerebral blood flow in the prefrontal cortex should be small.

In contrast, problem-appropriate knowledge which could be developed following the provision of problem-appropriate instruction in table use could enable successful table use for solving and calculating the correct answer to the problem given. If that occurs, we predicted that the cerebral blood flow in the prefrontal cortex would increase in order to execute the executive function of working memory, as previously explained. In other words, we predicted that differences in cerebral blood flow in the VLPFC and DLPFC regions may indicate differences in the general or problem-appropriate knowledge about table use that our participants possessed.

The following were the corresponding hypotheses we tested:

> H1. That, when prompted to construct tables during problem solving, the majority of students would be able to do so—hence, that table use would increase after providing a prompt for its use.

> H2. That correct answer rates following the prompt to use tables would evidence no change compared to previous performance.

> H3. That in the VLPFC and DLPFC regions of the brain, cerebral blood flow would decrease after the prompt for students to use tables in their problem solving.

> H4. That spontaneous table use would increase after the provision of problem-appropriate instruction in table use.

> H5. That correct answer rates would increase after the provision of problem-appropriate instruction in table use.

> H6. That in the VLPFC and DLPFC regions of the brain, cerebral blood flow would increase after the provision of problem-appropriate instruction in table use.

Materials and methods

Ethics approval

The conduct of this study was approved by the research ethics committee of the university to which all the authors belong, with the exception of the third author. All the students who participated in the study did so voluntarily and provided informed consent for their participation. In the case of minors, parental consent was also obtained.

Participants

The participants were 16 right-handed students from elementary school to undergraduate university levels (female = 6; mean age = 15.7 ± 2.9 years, from 10 to 19 years; laterality quotient = + 69.6 ± 22.5, Oldfield, 1971) recruited from students who were attending an after-school tutoring service in Japan. The decision on the wide age range of participants for the study was based on previous findings (Ayabe et al., 2021b). In that Ayabe et al. study, no significant age effect was found for scores in diagram use, although they conducted similar experiments. They provided instruction to improve diagram knowledge about tables in 40 students ranging in age from 10 to 20 years (15.0 ± 3.1 years old). The results showed no correlation between age and frequency of table use, and correct answer rates when solving the mathematical word problem.

We used G*Power (Faul et al., 2007) to estimate the minimum sample size for our within-participant design with three phases. That estimated that seven participants would be required to detect a statistically significant difference for the assumed large effect size based on the Ayabe et al. study ($f = 0.54$, alpha-level $p = 0.05$, power = 0.80). As the sample sizes of one group in previous fNIRS studies in mathematics and language were from 8 to 15 (see Soltanlou et al., 2018 for a review), we decided the necessary sample size to be 15. In the experiment, 20 students participated, but four missed some sessions. We used data from 16 students in the analyses after excluding the four.

To confirm the equivalence of three mathematical word problems that were used during the assessment sessions, the participants were randomly assigned to one of three groups according to the sequence with which they received those problems (Group A, $n = 6$; Group B, $n = 5$; Group C, $n = 5$; see **Table 1**).

Experimental design and procedure

A pre- and post-intervention design within participants, with four phases was used (see **Table 1** for an outline). In Phase 1 (Pretest 1), participants were asked to solve one of three mathematical word problems according to the group they were randomly assigned. Apart from being asked to solve the problem, no other instruction or information was provided. Phase 2 (Pretest 2) was identical to Phase 1 in that participants were again asked to solve a different but isomorphic problem. However, this time, they were explicitly asked to construct and use a diagram—more specifically, a table—in their attempts

TABLE 1 Outline of the experimental design and procedure.

Phase	Purpose	Task and requirement	fNIRS	Problems to solve		
				Group A	Group B	Group C
1	Pretest 1	Solve problem; no requirement	Yes	Time	Sweets	Parts
2	Pretest 2	Solve problem; use table	Yes	Sweets	Parts	Time
3	Intervention	Receive instruction; practice	No	–	–	–
4	Post-test	Solve problem; no requirement	Yes	Parts	Time	Sweets

at solving the problem. Phase 3 was the Intervention phase, during which the participants received instruction and practice in the construction and use of tables in mathematical word problems like those administered in the assessment sessions of this study. Finally, in Phase 4 (Post-test), the participants were administered a third problem to solve (different but again isomorphic to the previous two they had received). Like in Phase 1, they received no other instruction or requirement (i.e., it was up to them how to approach solving the problem, including whether to use a table or not).

During Phases 1, 2, and 4, brain activity data (cerebral blood flow) was also taken using fNIRS (functional near-infrared spectroscopy) equipment during the time the participants were solving the word problems. These phases were conducted in a laboratory to enable the use of the fNIRS equipment. We also measured the solving time, which is the time taken to solve each of the mathematical word problems. The solving time provides a measurement of the participants' duration of engagement in relation to their problem solving performance and cerebral blood flow. The instruction provided in Phase 3 was conducted in a different experimental room. Phases 1, 2, and 4 took approximately 20 min each (including attachment and detachment of the fNIRS equipment), while Phase 3 took 30 min. All procedures were conducted individually for each participant. For each participant, all sessions were held on the same day. The four phases were conducted in sequence, one after the other, with short rest breaks of about 5 min in between.

All the materials used in the experiment were collected immediately after each phase to ensure that they would have no influence on the participants' learning or problem solving performance outside of the sessions provided (e.g., that the participants would not look over and think about them during the short breaks).

Problems used

The problems used were all mathematical word problems in which tables were deemed to be helpful for solving. They all required the apprehension of the pattern of change in numbers described in the problem, and the prediction of a future quantity (see **Table 2**). All the problems were presented in sentences only and did not include expressions to explicitly induce the use of diagrams. Preliminary statistical analysis confirmed their isomorphism in terms of difficulty, as no significant differences between problems occurred in all phases. Here, isomorphic problems refer to problems that have different cover stories but have equivalent solving structures.

The problem stories were printed on the answer sheets provided to the students, so that it would be easier for them to read them as well as underline or mark any parts of the stories they want to. The question corresponding to the problem given was displayed on the computer screen. The text length of the mathematical word problems was about 110 characters in Japanese (about 57 words in English). They were deemed too complicated to be solved mentally. However, if the participants identified and then arranged the necessary numbers/quantities in sequence in an array or table according to the rules provided in the problem, they were expected to discover the pattern of change. That would then facilitate inductive reasoning and, as a result, the ability to predict or calculate the future quantity required. Similar problems are included in Japanese elementary school textbooks (e.g., Souma, 2020).

Experimental setup for problem solving and functional near-infrared spectroscopy measurement

After attaching the fNIRS holder to his/her head to measure brain activity (cerebral blood flow), each participant was seated in the laboratory (in the shield room constructed so that it is not affected by external electric or magnetic fields and does not leak to the outside.). A computer monitor was installed on a desk in front of the seat, and instructions and questions were displayed on that monitor (see **Figure 1**). The LCD monitor (21.5 inches) was 70 cm away from the participant's face.

The test procedures performed in Phases 1, 2, and 4 were all the same, but the problems provided according to group assignment were different (see **Table 1**). The time course was as follows. First, the participants were asked to rest with their eyes closed (30 s). After that, the problem answer sheet, which also contains the problem story was provided to the

TABLE 2 Problems used.

Name	Problem wording (translated from Japanese)
Time	*Studying and playing time*: A boy spends 150 min every day studying and playing games. If his study time is longer than playing games, he will play games 30 min longer the next day. Otherwise, he will study 45 min longer than the time spent playing games on the previous day. He studied for 50 min on day 1. (Question) How many minutes will he study for on day 365?
Sweets	*Sweets tree*: There is a sweets tree that bears 111 macaroons and eclairs every day. If the eclairs are more than the macaroons, it will bear 33 more macaroons the next day than the number of eclairs on the previous day. Otherwise, it will bear 22 fewer macaroons than the previous day. It bore 99 macaroons on day 1. (Question) How many macaroons will the tree bear on day 365?
Parts	*Automobile factory*: A factory manufactures automobile parts, packing 150 parts in each box, and carries as many boxes as possible to an assembly plant every day. If the number of parts left is more than 75, the factory will produce 180 parts the next day. Otherwise, it will produce 220 parts. The number of the parts left on day 1 was 130. (Question) How many parts will be left on day 365?

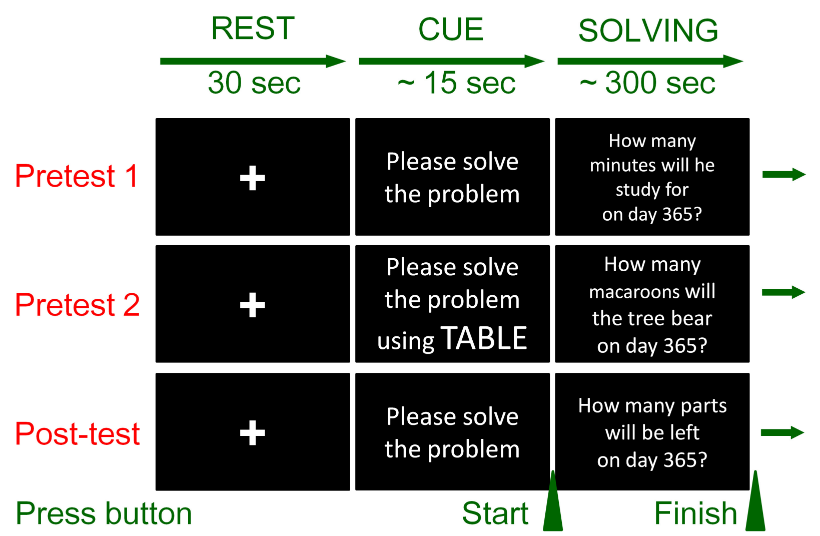


FIGURE 1
Test condition and procedure. Example of Group A (problems presented depends on the group; see Table 1). The procedures for all three test sessions were the same. Only in pretest 2 was the prompt to solve using the table given in the clue phase.

participants (see Table 2). The question for the problem was displayed on the computer monitor on the desk. After reading the problem story and the question, the participant could then start his/her attempt at answering the question (for which they had up to 5 min). Note that, only in Phase 2, instruction to "Please solve the problem using a TABLE" was displayed on the computer screen just before (i.e., above) the question for the problem. To ensure that the student would follow the instruction, the experimenter asked if the student had understood the display and only allowed the student to move on to the problem solving phase after receiving a YES response.

Instruction

In Phase 3, the intervention session, instruction to develop procedural and conditional knowledge for the use of tables was provided. This phase comprised two parts: instruction and practice. The third author and a teacher with no vested interest in the outcomes of this research provided the instructions and facilitated the practice session for each participant individually. To ensure procedural fidelity and consistency in the conduct of this intervention, PowerPoint slides and worksheets were used. The intervention was conducted in an experimental room different from the shielded room used in Phases 1, 2, and 4.

In the instruction part, how to determine the rules that govern change in the relevant numbers provided in a problem by arranging them in a table was demonstrated and explained. Then, how to inductively draw inferences and calculate the required amount based on the "rule" that has been apprehended was covered. Mathematical word problems, which were different but isomorphic in structure and requirements to those administered in Phases 1, 2, and 4, were used in these demonstrations and the subsequent practice. The instruction part took 10 min. In the practice part, which took 20 min, the participant was asked to solve two problems and was given feedback and advice as was required to ensure that he/she could correctly solve those problems by constructing and using tables (see **Table 3**).

Behavioral data acquisition

The presence/absence of a table (Table included/used: Yes/No) created by the participants, the correctness of all the numbers in a table students provided (Correct numbers: Yes: 1, No: 0), the correctness of the formulas to find the answer after inferring using the table (Correct formulas: Yes: 1, No: 0), and the answer they provided (Correct answer provided: Yes/No) were examined. Two teachers with no vested interest in the research outcomes undertook this scoring independently. The values of the kappa coefficients showing the agreement between the inter-raters (i.e., these two teachers) were 0.95 in Table use, 1.00 in the correctness of numbers, formulas, and answers, indicating almost perfect and perfect concordance, respectively (Cohen, 1960; Landis and Koch, 1977).

Functional near-infrared spectroscopy data acquisition

An fNIRS device was used to determine cerebral blood flow in the frontal lobe during solving of the mathematical word problems. For near-infrared spectroscopy, a multi-channel spectrometer operating at 780-, 805-, and 830-nm wavelengths (Foire-3000, Shimadzu, Japan) was used to measure temporal changes in concentrations of oxygenated hemoglobin (oxy-Hb), deoxygenated hemoglobin, and total hemoglobin. We used only oxy-Hb changes for analysis because oxy-Hb corresponds most with the blood oxygenation level-dependent signal of fNIRS (Strangman et al., 2002).

TABLE 3 Instruction procedure.

Teaching table use strategy knowledge	
1	**Declarative knowledge: Knowing that** A table is a visual representation that allows arrangement of numbers/quantities in lines or rows for each item and for organizing them compactly for easy viewing
2	**Conditional knowledge: Knowing when and why** To use a table to find the change rule when solving problems with varying numbers/quantities
3	**Procedural knowledge: Knowing how** 1. Create a table focusing on the number/quantity required to solve the problem. 2. Find the change rule, such as the repetition period, from the sequence in the table. 3. Inductively infer general laws from the change rules. 4. Calculate the answer using the general formula.

Diagram use instruction		**Example of parts problem**
Constructing stage		
1	Read the problem carefully	
2	Underline necessary numbers and conditions	
3	Identify changing number (variables)	Day, the number of stock before (B) and after shipping (A)
4	Create a frame for the table, if necessary	
5	Write the numbers given in the problem into the table	130 as the number of stock (B) on Day 1
6	Calculate the numbers of stock (B, A) on Day 2 according to the problem condition	(310, 10)
7	Similarly, calculate the numbers of stock (B, A) after Day 3	(235, 85), (265, 115), (295, 145), (325, 25), (250, 100), (280, 130), (310, 10)
Inferring and computing stage		
8	Infer the change rule of the numbers of stock (B, A)	If the same value appears on the table, for example on Days 1 and 8, then it may signal that one periodic sequence is complete and is being repeated.
9	Find the periodic length	7 (days)
10	Calculate the remainder by dividing 365 by the periodic length	365 ÷ 7 = 52 remainder 1
11	Find the number in the order of remainder in the periodic sequence	130 (the 1st number of the periodic sequence)

The example above was in the case of the Parts problem (see Table 2). The experimenter taught students how to construct and use tables for solving the problem according to the above instructions. After the instructions, the experimenter had the students practice two more isomorphic problems. In the instruction and practice sessions, we used isomorphic problems different from those used in Pretest 1, Pretest 2, and Post-test.

We used a 3 × 9 array of 27 probes (fNIRS Folder for forehead measurement 551-07600-01; 14 light-emitting and 13 detecting probes), which enabled measuring in 42 channels (see **Figure 2**). Following the international 10–20 system (Sharbrough, 1991), we fitted 19 ch. and 24 ch. to F8 and F7, respectively. This fitting placed 38 ch. around Fp1 and 39 ch. around Fp2. The sampling time was 220 msec. In order to minimize artifacts, we asked participants to, as much as possible, limit their bodily movements to only those that are necessary to execute the task administered (i.e., solving the problem).

Before analysis, the fNIRS data were filtered through a low pass filter at 0.5 Hz and fifth order. Since cerebral blood flow changes several seconds later than neural activity, fNIRS data for the initial 10 s of stimuli onset in rest and solving phases were removed from the analysis. We subtracted the average oxy-Hb during the rest phase (30 s) from the average during the solving phase for each phase (e.g., Pretest 1—rest) and defined this relative oxy-Hb change as activation of the pre-frontal cortex (PFC) (Δoxy-Hb). After that, we subtracted the Δoxy-Hb in the two tests (Pretest 2—Pretest 1, Post-test—Pretest 2, Post-test—Pretest 1) and calculated the changes between the solving phases. Note that the rest phase immediately prior to each test session was set as the baseline for each test.

We estimated the anatomical data of the fNIRS channel using a three-dimensional digitizer and mapped the anatomical data on the image of the brain using a mapping software (Fusion, Shimadzu, Japan). As previously noted, the area of interest in the present study was the dorsolateral prefrontal cortex (DLPFC) and the ventrolateral prefrontal cortex (VLPFC). According to the international 10–20 system, the target channels based on this fNIRS probe placement were 3 ch. for right-DLPFC, 6 ch. for left-DLPFC, 36 ch. for right-VLPFC, and 41 ch. for left-VLPFC (Himichi and Nomura, 2015; Jin et al., 2019).

Analysis

We used three kinds of analysis to examine the dependent measures (behavioral and physiological data) depending on the variable type. First, we used Cochran's Q test, a non-parametric statistical test, to analyze the data about the students' table use and the correctness of their answers, as these were binary data (0 or 1). For pairwise comparisons of simple main effects, we used McNemar's test as it is likewise suitable for detecting differences in dichotomous dependent variables. The null hypotheses were set against each effect of prompt for table use (before Pretest 2) and problem-appropriate instruction on table use (before Post-test). Second, we used paired t-tests to analyze the students' cerebral blood flow before and after the prompt and the instruction. In order to understand variations in blood flow in the entire frontal lobe, we performed an analysis on all 42 channels measured. However, further analysis focused on the four channels above (the ROIs).

Results

Table 4 shows the results of the acquired behavioral data. And **Figure 3** shows an example of tables drawn by the students for Pretest 1, Pretest 2, and Post-test. In Pretest 1, most students wrote numbers, calculation formulas, and calculations on paper, like in **Figure 3**. Four students drew a table, and another drew an illustration. In Pretest 2, 11 students drew tables, one drew an illustration, one drew a flow chart, and the other three only used calculation formulas. In Post-test, all 16 students drew (nearly)

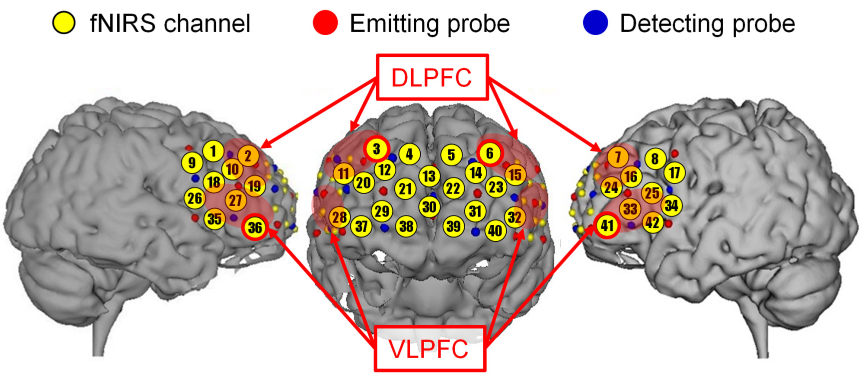


FIGURE 2
Anatomical space of each functional near-infrared spectroscopy (fNIRS) channel. Regions of interest (ROIs) are circled in red. Upper left circle, right DLPFC (3 ch.); upper right circle, left DLPFC (6 ch.); lower left circle, right VLPFC (36 ch.); lower right circle, left VLPFC (41 ch.); Red shadings are the estimated anatomical regions (right DLPFC, 2, 3, 10, 11, and 19 ch.; left DLPFC, 6, 7, 15, 16, and 24 ch.; right VLPFC, 18, 27, 28, 35, and 36 ch.; left VLPFC, 25, 32, 33, 41, and 42 ch.; e.g., Yamaya et al., 2021).

TABLE 4 Table use, numbers, formulas, and correct answer.

	Table use		Numbers		Formulas		Correct answer	
	Use	Rate	Correct	Rate	Correct	Rate	Correct	Rate
Pretest 1	4	0.25	1	0.06	0	0.00	0	0.00
Pretest 2	11	0.69	4	0.25	1	0.06	1	0.06
Post-test	16	1.00	7	0.44	6	0.38	6	0.38

N = 16.
Table use: the presence of a table students provided (Yes: 1, No: 0); Numbers: the correctness of all the numbers in the table (Yes: 1, No: 0); Formulas: the correctness of the formulas for finding the answer by inferring using the table (Yes: 1, No: 0); Correct answer: the correctness of the answer students provided (Yes: 1, No: 0).

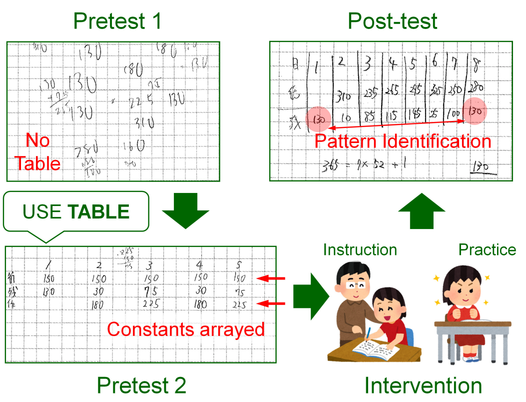


FIGURE 3
Examples of TABLES drawn by students.

complete tables. Note that, for any test, it does not matter if the values in the table are correct or not.

Did table use increase after the provision of a prompt for its use?

One of the four students who constructed and used a table in Pretest 1 did not construct a table in Pretest 2 despite being prompted to use a table. However, eight students followed the prompt and constructed a table for the first time in the experiment.

A comparison of participant table use in Pretest 1 and Pretest 2 revealed that table use significantly increased following the prompt to construct and use it during problem solving. This can be seen in **Figure 4**, which shows Table use as a function of the test phases. The percentages of answer sheets that included a table (represented by the red bars) before and after the prompt (i.e., at Pretest 1 and at Pretest 2) are shown. In Pretest 1, most of the participants tried to solve the mathematical word problem by calculation alone, without using tables. In Pretest 2, as expected, after the provision of the prompt to use a table, table use increased.

The analysis showed a significant test phase effect for table use, $Q_{(2)} = 16.77$, $p < 0.001$ [where $Q_{(df)}$ = Cochran's Q value with the corresponding degrees of freedom]. Moreover, a significant difference was found in table use between the test immediately before and after the prompt for table use [Pretest 1 vs. Pretest 2: $\chi^2_{(1)} = 5.44$, $p < 0.05$, where $\chi^2_{(df)}$ = McNemar's χ^2 with the corresponding degrees of freedom].

This result provides full support for the first hypothesis. Note, however, that in Pretest 2, five students did not construct a table despite being instructed to do so. As noted in the "Materials and methods" section, the experimenter asked all students to use a table by displaying the prompt on the monitor and soliciting verbal confirmation of the student's understanding of the instruction. Also, all students wrote something on their answer sheet (i.e., none of them failed to write anything).

Did correct answer rates increase after the prompt for table use?

No students answered the problem correctly in Pretest 1. Only one student answered correctly in Pretest 2. That student constructed a table in both tests (i.e., with and without the prompt), failed to correctly solve the problem in Pretest 1, but succeeded in Pretest 2.

A comparison of participants' correct answer rates in Pretest 1 and Pretest 2 revealed no improvement in their correct answer rates. **Figure 4** shows table use as a function of the test phases. The blue bars in **Figure 4** show that no correct answers were produced in Pretest 1 and only a very slight increase occurred at Pretest 2. The statistical analysis confirmed this—that is, despite a significant test phase effect for correct answer rates, $Q_{(2)} = 10.33$, $p < 0.01$ [where $Q_{(df)}$ = Cochran's Q value with the corresponding degrees of freedom], there was no significant difference between the correct answer rates immediately before and after the prompt to use tables [Pretest 1 vs. Pretest 2: $\chi^2_{(1)} = 1.00$, $p \geq 0.1$].

This result provides full support for the second hypothesis. In other words, despite using tables in attempts to solve the problem given during Pretest 2, there was no significant improvement in the participants' ability to produce correct solutions.

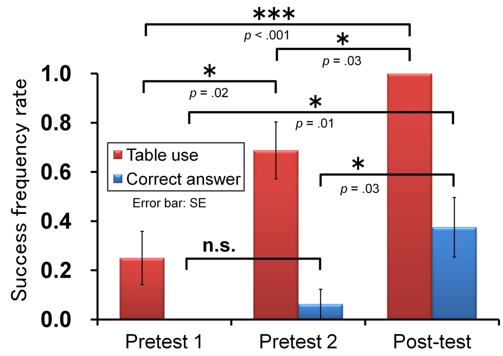


FIGURE 4
Changes in table use and correct answer rate (*N* = 16). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

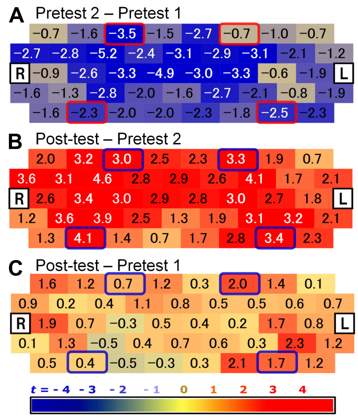


FIGURE 5
Changes in cerebral blood flow. The numbers on the figure are t values for changes in the average cerebral blood flow concentration per unit time between test phases; **(A)** Pretest 2—Pretest 1; **(B)** Post-test—Pretest 2; **(C)** Post-test—Pretest 1; (R) right, (L) left; *N* = 16. ROIs are circled. Upper left circle, right DLPFC (3 ch.); upper right circle, left DLPFC (6 ch.); lower left circle, right VLPFC (36 ch.); lower right circle, left VLPFC (41 ch.) (see Figure 2).

Did the cerebral blood flow in the dorsolateral prefrontal cortex and ventrolateral prefrontal cortex regions of the brain decrease after the prompt for table use?

As explained in the Introduction section, the students were entirely free to solve the mathematical word problem in Pretest 1 using any approach they liked. However, in Pretest 2, they were restricted in their answering approach by being asked to use a table. Eight students changed their approach from calculating only to using tables, and five students were not able to follow the prompts. Based on these, we compared the cerebral blood flow data in Pretest 1 with those in Pretest 2 (see Figure 5).

Significant decreases in cerebral blood flow were found in the right DLPFC area as well as the left and right VLPFC areas of the brain following the prompt for table use. However, no decrease was found in the left DLPFC region. These are portrayed in Figure 5A, which shows changes in cerebral blood flow while solving the mathematical word problem in Pretest 2 compared to Pretest 1. The color scale indicates the changes on the measured channel from red (increase) to blue (decrease). Circles on the figure indicate the regions of interest (ROIs) in the brain area. That is, the upper left circle represents the right DLPFC (3 ch.), the upper right circle represents the left DLPFC (6 ch.), the lower left circle represents the right VLPFC (36 ch), and the lower right circle represents the left VLPFC (41 ch.). As Figure 5A shows, the cerebral blood flow other than in the left DLPFC, more specifically in the right DLPFC area and left/right VLPFC areas, decreased when comparing Pretest 2 to Pretest 1.

The numbers in Figure 5A indicate the t-test values in comparing the blood flow in Pretest 2 to that in Pretest 1. Significant differences in cerebral blood flow were found in the right DLPFC [$t_{(15)} = -3.47$, $p < 0.01$], right VLPFC [$t_{(15)} = -2.26$, $p < 0.05$] and left VLPFC [$t_{(15)} = -2.45$, $p < 0.05$] areas. However, no significant difference in cerebral blood flow was found in left DLPFC area [$t_{(15)} = -0.69$, $p \geq 0.1$].

These results provide partial support for the third hypothesis that, despite using tables to solve the problem during Pretest 2, rather than an increase in blood flow in these ROIs, there was instead a tendence for blood flow to decrease.

Did spontaneous table use increase after the provision of problem-appropriate instruction in table use?

In Pretest 2, 11 students constructed a table following the prompt, but in Post-test, all 16 students constructed a table without being prompted (see Table 4). A comparison of participant table use in Pretest 2 and Post-test revealed that Table use significantly increased following the provision of problem-appropriate instruction in table use. This is shown by the red bars in Figure 4.

The statistical analysis revealed a significant test phase effect for Table use, $Q_{(2)} = 16.77$, $p < 0.001$. Moreover, significant difference was found when comparing Table use between Pretest 2 and Post-test (i.e., before and after instruction had been provided), $\chi^2_{(1)} = 5.00$, $p < 0.05$.

These results provide full support for the fourth hypothesis. In other words, there is evidence indicating that provision of problem-appropriate instruction in table use led to improvements in participants' ability to use tables spontaneously to solve the type of mathematical word problems.

Note that in Post-test, the students were not prompted to use the table to solve the problem.

Did correct answer rates increase after the provision of problem-appropriate instruction in table use?

Only one student succeeded in deriving the correct answer in Pretest 2, but six students succeeded in the Post-test. The only student who succeeded in Pretest 2 also succeeded in Post-test. Five students failed to obtain the correct answer in Pretest 2 despite constructing tables but succeeded for the first time in the Post-test (see **Table 4**).

There was a significant increase in the correct answers produced by participants at Post-test compared to Pretest 2. This is shown by the blue bars in **Figure 4**. The statistical analysis revealed a significant test phase effect for the correct answer rates [$Q_{(2)} = 10.33$, $p < 0.01$]. Moreover, significant difference was found when comparing correct answer rates between Pretest 2 and Post-test (i.e., before and after instruction had been provided), $\chi^2{}_{(1)} = 5.00$, $p < 0.05$.

These results provide full support for the fifth hypothesis. In other words, there is evidence indicating that provision of problem-appropriate instruction in table use led to improvements in participants' ability to solve the type of mathematical word problems that would benefit from the use of tables (i.e., those that require working out the pattern and rule of change in amounts). There were, however, ten students who failed to obtain the correct answer in the Post-test, and they constructed nearly complete tables. A close examination of their answer sheets showed that although they all undertook the appropriate approach in attempting to solve the problem, they made mistakes in the arithmetic procedure to determine or infer the values/numbers that make up the tables.

Did cerebral blood flow in the dorsolateral prefrontal cortex and ventrolateral prefrontal cortex regions of the brain increase after the provision of problem-appropriate instruction in table use?

In Pretest 2, we instructed students to use a table, but most of them could not create and use a table as the problem demanded. As described before, all students used a table spontaneously in Post-test. The tables they drew were nearly complete ones (see **Figure 3**), so they likely engaged in similar cognitive activities regardless of their success or failure in correctly solving the problem. Based on these, we compared the cerebral blood flow while solving the mathematical word problem in Pretest 2 with that in the Post-test.

There was a significant increase in cerebral blood flow to the left and right DLPFC and VLPFC regions of the brain at Post-test compared to Pretest 2. This is shown in **Figure 5B**: pay particular attention to the circled ROIs. The numbers in **Figure 5B** indicate the *t*-values on each of the fNIRS channels. When comparing cerebral blood flow readings between Post-test and Pretest 2, the analysis found significant differences in blood flow in the right DLPFC [$t_{(15)} = 3.00$, $p < 0.01$], the left DLPFC [$t_{(15)} = 3.29$, $p < 0.01$], the right VLPFC [$t_{(15)} = 4.13$, $p < 0.01$], and the left VLPFC [$t_{(15)} = 3.35$, $p < 0.01$].

These results provide full support for the sixth hypothesis: following problem-appropriate instruction in table use, there was a significant increase in blood flow to the DLPFC and VLPFC regions of the brain when participants were working on the problem they were given, suggesting active engagement of working memory functions noted earlier.

Finally, although not directly related to the hypothesis of this study, the results of comparing brain activity in Pretest 1 and the Post-test are shown in **Figure 4C**. No significant difference was found in any ROI. Apart from the ROI, a significant difference was seen on only one channel (33 ch.; $t_{(15)} = 2.32$, $p = 0.04$). Therefore, this result shows the brain activities in Pretest 1 and Post-test were comparable if considering the multiple tests.

Means of cerebral blood flow and solving time by success or failure

For a more detailed analysis, we considered it necessary to examine the differences in brain activity between students who succeeded in the Post-test (6 students) and those who failed (10 students). As mentioned above, we speculated that they engaged in similar cognitive activities/efforts regardless of their success or failure outcome in solving the problem. However, those small sample sizes made statistical validation useless. We report only aggregated results here. **Tables 5**, **6** show the average cerebral blood flow in all channels and ROIs (four channels), respectively. We also considered solving time to be an essential factor. If students engaged in similar cognitive activities/efforts regardless of success or failure, there should be no difference in the solving time. However, for the same reason above, it was

TABLE 5 Cerebral blood flow (prefrontal cortex).

	All		Success ($N = 6$)		Failure ($N = 10$)	
	Mean	SD	Mean	SD	Mean	SD
Pretest 1	0.0181	0.0182	0.0178	0.0224	0.0184	0.0165
Pretest 2	0.0014	0.0178	0.0014	0.0238	0.0014	0.0145
Post-test	0.0308	0.0268	0.0462	0.0320	0.0154	0.0152

$N = 16$.
Mean cerebral blood flow and the standard deviation for all 42 channels (mM·mm).

TABLE 6 Cerebral blood flow (ROI).

	All		Success (N = 6)		Failure (N = 10)	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Pretest 1	0.0914	0.0951	0.1089	0.1229	0.0739	0.0785
Pretest 2	0.0168	0.0613	0.0358	0.0815	−0.0022	0.0445
Post-test	0.1664	0.1563	0.2334	0.2164	0.0993	0.0850

N = 16.
Mean cerebral blood flow and the standard deviation for the four ROI channels (mM·mm).

TABLE 7 Solving time.

	All		Success (N = 6)		Failure (N = 10)	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Pretest 1	203.6	93.6	219.7	78.7	187.6	104.4
Pretest 2	230.6	81.8	258.8	49.4	202.4	92.1
Post-test	238.7	61.7	226.3	57.6	251.1	65.2

N = 16.
Mean solving time and the standard deviation (sec).

difficult for us to report statistical analysis results, so **Table 7** shows only the aggregate results of the solving time.

Discussion

Summary of the main finding and interpretation of results

In the present study, we were able to obtain performance/behavioral and neurophysiological evidence of the effect of providing problem-appropriate instruction, which develops students' ability to construct and use diagrams for solving mathematical word problems. As far as we know, this is the first study to report cerebral blood flow changes as a consequence of developing strategy use knowledge in problem solving. Our results showed that:

(1) If students were required to use diagrams (tables), they could construct them. However, such construction rarely led to successful problem solving.

(2) In that case, cerebral blood flow to regions of the brain related to working memory functions (i.e., the VLPFC and DLPFC areas) decreased compared to when students were not required to use tables in problem solving.

(3) After problem-appropriate instruction had been provided, both the students' spontaneous use of diagrams (i.e., without prompting) and their production of correct answers increased.

(4) In that case, cerebral blood flow to the regions of the brain associated with working memory functions increased compared to measurements prior to the provision of instruction.

Evaluation of methods and assumptions

In this study, we experimentally verified the distinction between two kinds of diagram use knowledge (i.e., general and problem-appropriate) in the way they affect problem solving—in terms of both performance and brain response. We were able to confirm that problem-appropriate instruction can develop problem-appropriate knowledge that can address prevalent student deficits in the production and use of diagrams (Zahner and Corter, 2010; Van Garderen et al., 2013; Reuter et al., 2015). This finding may help contribute to overcome the repeatedly reported the utilization deficiency of diagrams (Uesaka and Manalo, 2006; Uesaka et al., 2010; Ayabe and Manalo, 2018).

Furthermore, we were able to demonstrate the usefulness of fNIRS as a tool for establishing neurophysiological evidence for strategy use acquisition—in near normal teaching and learning settings (i.e., participants were able to behave as they normally would in problem solving, which would have been difficult to do with the use of more movement-restrictive equipment such as fMRI and EEG; Soltanlou et al., 2018).

Finally, the mathematical word problems designed for this study helped to control table use knowledge effectively. Suppose the problem is not difficult to solve without tables or is easily solved by external representations other than tables. In that case, the assumption that table use is appropriate does not hold (e.g., Reuter et al., 2015). If those mathematical word problems were such, it would not have distinguished the three kinds of knowledge (Paris et al., 1983) for mastering table use strategies. In this study, all students failed in Pretest 1, where they were allowed to solve freely, and in the Post-test after table use knowledge instruction, the correct answer rate improved as table use increased. Therefore, the mathematical word problems used in this study can be taken as having worked to identify the components of table use knowledge.

Answers to questions raised

We had three main questions we were trying to answer in this research. The first of those was whether students would be able to construct tables when asked to do so for solving mathematical word problems (for which the use tables would be considered appropriate) (H1), and whether such demand would lead to increases in correct answers (H2). First of all, researchers have often reported that students do not construct diagrams spontaneously, even though they know diagrams help solve mathematical word problems (e.g., Uesaka et al., 2007). Only four students (25%) constructed a table in the Pretest in this experiment. This result confirmed the problem of lack of spontaneity in diagram use. Based on these, this question was generated in consideration of an advice often given by teachers in mathematics classrooms (i.e., "Use diagrams to solve this math word problem!"). As expected, asking students to use a table led to an increase in the use of tables during problem solving (from 4 to 11 students), but the correct answer rates did not improve (from 0 to 1 student). This result is congruent with those of previous studies that students, when encouraged to use diagrams, can construct diagrams—but that, without adequate problem-appropriate instruction, they are often unable to effectively use the diagrams for the task given (Uesaka and Manalo, 2006; Uesaka et al., 2010; Manalo and Uesaka, 2016).

As mentioned earlier, all students would know what a table is (Ayabe et al., 2021a). We also carefully gave them instructions that they needed to use a table for solving the problem given (monitor display and verbal confirmation). Therefore, the five students who did not construct a table probably gave up on constructing one because they did not know how to use it. Also, of the 11 students who constructed a table according to the instructions, only one student succeeded in Pretest 2. The rest (10 students) failed to solve the problem even though prompted to use a table.

A simpler mathematical word problem may be possible to solve by drawing items described in the problem and constructing a schema (e.g., Boonen et al., 2013; Daroczy et al., 2015): such diagram construction can be undertaken with the use of general knowledge about diagrams as visual representations. However, such general knowledge is insufficient for solving more difficult word problems, which require problem-appropriate knowledge for constructing and using the appropriate diagram so that it would not only visually represent the problem but also enable its actual solving (e.g., in the case of tables, it should enable the apprehension of the pattern of recurring change in numerical values; cf. Duval, 1999, 2006). This point highlights the importance of improving teachers' teaching methods for diagram use strategy. The advice to "use a table to solve this problem," which is the extent of explicit instruction many teachers provide, is far from adequate as it only provides declarative knowledge (i.e., it only answers the questions of: *What* strategy should I use? cf. Paris et al., 1983). The limited value of such advice was clearly demonstrated by our result of students being able to construct tables but failing to use them to work out the required answers.

The second question was whether the provision of problem-appropriate instruction about table use would further increase students' spontaneous use of tables during problem solving (H4) and improve their ability to produce correct answers (H5). More specifically, all 16 students constructed a table spontaneously and completely in the Post-test. Six of them succeeded in correctly solving the problem. As mentioned above, the cause of the failure appeared to have been calculation errors (**Table 4**). Therefore, it indicates that the instruction and practice on table use provided in the intervention session helped them overcome the lack of spontaneity and to acquire knowledge about table use. This result is consistent with previous studies that have provided similar interventions (Ayabe and Manalo, 2018; Ayabe et al., 2021b).

The results at Post-test showed that both table use and correct answers increased. The problem-appropriate knowledge about table use provided during instruction students to not only construct a more useful table for the problem they were given (paying attention and extracting relevant numerical and other details from the text of the problem itself), but also to visually apprehend patterns of number changes, to inductively infer the rules that govern such patterns, and to undertake more efficient calculations of the answer required. The crucial point is that after teaching them that knowledge, the students possessed the necessary procedural and conditional knowledge appropriate for solving the type of problem given. Without problem-appropriate instruction, the table can be drawn, but it will unlikely lead to the production of the correct solution.

The third question was whether neurophysiological measurements (using fNIRS) could detect differences between the possession of general and problem-appropriate knowledge about table use. This study focused on the relationship between the working memory system and frontal cerebral blood flow (Wijeakumar et al., 2017; Soltanlou et al., 2018). At Pretest 1, the students would have been free to solve the problem given in any way they wanted, including just trying to numerically compute the answer and drawing pictures or illustrations. Brain activity measurement at this phase might reflect the state of natural mental effort and cognitive activity. At Pretest 2, cerebral blood flow was found to be lower overall than at Pretest 1 (H3). The reason may have been due to deficiencies in the students' diagram use strategy (Bjorklund et al., 1997). In other words, the requirement to use a table might have turned into a stumbling block because, although most of the students would have known how to construct a basic table, many proved not to know how to construct and use a table to solve the problem they were given (cf. Uesaka and Manalo, 2006; Uesaka et al., 2010). The requirement to use an appropriate diagram (table) that they did not have sufficient knowledge about, could have activated the ACC area of the brain instead to try to resolve the problem (Botvinick et al., 1999; Cao et al., 2016), or the mPFC area to search for the way to use the table (Euston et al., 2012; Kurczek et al., 2015; Yonelinas et al., 2019). The net effect would have been the observed decreases in blood flow to the DLPFC and VLPFC regions of the brain. As previously noted, fNIRS would not have captured the activations in the deep PFC region (Fishburn et al., 2014), only the blood flow reductions in the DLPFC and VLPFC regions.

Finally, at Post-test, the cerebral blood flows became higher than those in Pretest 2 (H6). The activation of DLPFC and VLPFC would suggest that the working memory system is related to the use of problem-appropriate knowledge about table use (cf. Corbetta and Shulman, 2002; Petrides, 2005; O'Reilly, 2010), facilitating the execution of solving procedures, drawing of inferences, and more efficient computation (cf. Larkin and

Simon, 1987; De Smedt et al., 2011; Giardino, 2017). In addition, the goal-oriented thinking process may activate PFC (Curtis and D'Esposito, 2003).

Working memory demands could explain the reason for no difference in activity in the prefrontal cortex between Pretest 1 and Post-test. In other words, whether diagrams were used or not, those working memory demands were the same. Therefore, we can interpret that the table use improved problem-solving performance, even though no differences were found in brain activity.

Significance of findings

An important educational implication of the present study is that the problem-appropriate declarative, conditional, and procedural knowledge pertaining to strategy use (Paris et al., 1983) needs to be directly addressed through problem-appropriate strategy/diagram use instruction. Most teachers would show and tell students to use diagrams to solve mathematical word problems. However, such general and largely superficial instructions may be limited to just imparting declarative knowledge. The teachers may believe that the student would be able to solve problems if they provide a careful explanation, including the conditions for when to use the diagram. However, if the student cannot connect the declarative and conditional knowledge they have gained with the necessary procedural knowledge, it is likely that the student will fail in correctly using a diagram to solve the problem. To effectively connect and integrate declarative, procedural, and conditional knowledge for strategy use, problem-appropriate instruction on the strategy to use (in this case, tables for solving particular types of mathematical word problems) is necessary.

Limitation and future research

One limitation of this study is the small sample size ($N = 16$). In the future, it is necessary to confirm the reliability and validity of the findings obtained here on larger numbers of participants. Another limitation is that this study focused only on one particular type of mathematical word problems and the corresponding kind of diagram (i.e., table) deemed effective for solving it. Most diagrams have domain-specific functions, and their inference types are abundant (Ainsworth, 1999; Arcavi, 2003; Ayabe and Manalo, 2018). In order to contribute to improving students' diagram competencies, follow-up research is needed to cover the wider range of problem types and corresponding kinds of diagrams. Diagrams are an interactive and powerful thinking tool. Therefore, a better understanding of the correspondences between the functions that different kinds of diagrams can serve in mathematical word problem solving and the range of problem types that can be encountered, not only in the classroom but also in real life, can equip students with a powerful thinking tool for the duration of their formal education and the rest of their lives.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Research Ethics Committee Kyoto University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the minor(s)' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

Author contributions

HA and EM contributed to conception and design of the study. NH, HF, MN, and HA organized the experiment. HA performed the statistical analysis and wrote the first draft of the manuscript. EM, HF, and HA wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

HA's work on this research was supported by Grant-in-aid for JSPS (Japan Society for the Promotion of Science) Fellows (20J23507). EM's work on this research was supported by a grant-in-aid (20K20516) received from JSPS.

Acknowledgments

The authors express their gratitude to the students and teachers of Achieve Academy, Gifu, Japan, for their assistance in collecting and analyzing the data.

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

Ainsworth, S. (1999). The functions of multiple representations. *Comput. Educ.* 33, 131–152. doi: 10.1016/S0360-1315(99)00029-9

Ainsworth, S., Prain, V., and Tytler, R. (2011). Drawing to learn in science. *Science* 333, 1096–1097. doi: 10.1126/science.1204153

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educ. Stud. Math.* 52, 215–241. doi: 10.1023/A:1024312321077

Ayabe, H., and Manalo, E. (2018). "Can spontaneous diagram use be promoted in math word problem solving?," in *Diagrammatic representation and inference diagrqms 2018, lecture notes in computer science*, Vol. 10871, eds P. Chapman, G. Stapleton, A. Moktefi, S. Perez-Kriz, and F. Bellucci Cham: Springer International Publishing, 817–820. doi: 10.1007/978-3-319-91376-6

Ayabe, H., Manalo, E., Fukuda, M., and Sadato, N. (2021a). "What diagrams are considered useful for solving mathematical word problems in japan?," in *Proceedings of the diagrammatic representation and inference. Diagrams 2021. Lecture notes in computer science,* Vol. 12909, eds A. Basu, G. Stapleton, S. Linker, C. Legg, E. Manalo, and P. Viana (Cham: Springer), 79–83. doi: 10.1007/978-3-030-86062-2_8

Ayabe, H., Manalo, E., and Hanaki, N. (2021b). The effect of instruction on the domain specificity of diagrams based on the functional solution structure of mathematics word problems. *Jpn. J. Educ. Technol.* 45, 17–20. doi: 10.15077/jjet.S45011

Baddeley, A. (1992). Working Memory. *Science* 255, 556–559. doi: 10.1126/science.1736359

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2

Biondi, A., Santoro, V., Viana, P. F., Laiou, P., Pal, D. K., Bruno, E., et al. (2022). Noninvasive mobile EEG as a tool for seizure monitoring and management: A systematic review. *Epilepsia* 63, 1041–1063. doi: 10.1111/epi.17220

Bjorklund, D. F., Miller, P. H., Coyle, T. R., and Slawinski, J. L. (1997). Instructing children to use memory strategies: evidence of utilization deficiencies in memory training studies. *Dev. Rev.* 17, 411–441. doi: 10.1006/drev.1997.0440

Boisgueheneuc, F. D., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., et al. (2006). Functions of the left superior frontal gyrus in humans: a lesion study. *Brain* 129, 3315–3328. doi: 10.1093/brain/awl244

Boonen, A. J. H., de Koning, B. B., Jolles, J., and van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Front. Psychol.* 7:191. doi: 10.3389/fpsyg.2016.00191

Boonen, A. J. H., van der Schoot, M., van Wesel, F., de Vries, M. H., and Jolles, J. (2013). What underlies successful word problem solving? A path analysis in sixth grade students. *Contemp. Educ. Psychol.* 38, 271–279. doi: 10.1016/j.cedpsych.2013.05.001

Boonen, A. J. H., van Wesel, F., Jolles, J., and van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *Int. J. Educ. Res.* 68, 15–26. doi: 10.1016/j.ijer.2014.08.001

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., and Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* 402, 179–181. doi: 10.1038/46035

Brown, A. L., Canpione, J. C., and Day, J. D. (1981). Learning to learn: on training students to learn from texts. *Educ. Res.* 10, 14–21. doi: 10.3102/0013189X010002014

Cao, B., Li, W., Li, F., and Li, H. (2016). Dissociable roles of medial and lateral PFC in rule learning. *Brain Behav.* 6:e00551. doi: 10.1002/brb3.551

Cleveland, W. S., and McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Am. Stat. Assoc.* 79, 531–554. doi: 10.1080/01621459.1984.10478080

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

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., and Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage* 54, 2808–2821. doi: 10.1016/j.neuroimage.2010.10.069

Curtis, C. E., and D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423. doi: 10.1016/S1364-6613(03)00197-9

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., and Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature* 378, 279–281. doi: 10.1038/378279a0

Daroczy, G., Wolska, M., Meurers, W. D., and Nuerk, H.-C. (2015). Word problems: a review of linguistic and numerical factors contributing to their difficulty. *Front. Psychol.* 6:348. doi: 10.3389/fpsyg.2015.00348

De Smedt, B., Holloway, I. D., and Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. *NeuroImage* 57, 771–781. doi: 10.1016/j.neuroimage.2010.12.037

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science* 284, 970–974. doi: 10.1126/science.284.5416.970

Duval, R. (1999). "Representation, vision and visualization: Cognitive functions in mathematical thinking. Basic issues for learning," in *Proceedings of the Twenty First Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, Vol. 1, Cuernavaca, 3–26.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educ. Stud. Math.* 61, 103–131. doi: 10.1007/s10649-006-0400-z

Elia, I., van den Heuvel-Panhuizen, M., and Kolovou, A. (2009). Exploring strategy use and strategy flexibility in non-routine problem solving by primary school high achievers in mathematics. *ZDM Int. J. Math. Educ.* 41, 605–618. doi: 10.1007/s11858-009-0184-6

Euston, D. R., Gruber, A. J., and McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron* 76, 1057–1070. doi: 10.1016/j.neuron.2012.12.002

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fishburn, F. A., Norr, M. E., Medvedev, A. V., and Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Front. Hum. Neurosci.* 8:76. doi: 10.3389/fnhum.2014.00076

García, A. I., Jiménez, J. E., and Hess, S. (2006). Solving arithmetic word problems. *J. Learn. Disabil.* 39, 270–281. doi: 10.1177/00222194060390030601

Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Rev. Educ. Res.* 60:517. doi: 10.2307/1170504

Giardino, V. (2017). "Diagrammatic Reasoning in Mathematics," in *Springer Handbook of Model-Based Science. Springer Handbooks*, eds L. Magnani and T. Bertolotti Berlin: Springer, 499–522. doi: 10.1007/978-3-319-30526-4_22

Hegarty, M., and Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *J. Educ. Psychol.* 91, 684–689. doi: 10.1037/0022-0663.91.4.684

Hegarty, M., Mayer, R. E., and Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *J. Educ. Psychol.* 84, 76–84. doi: 10.1037/0022-0663.84.1.76

Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *J. Res. Math. Educ.* 23:242. doi: 10.2307/749120

Himichi, T., and Nomura, M. (2015). Modulation of empathy in the left ventrolateral prefrontal cortex facilitates altruistic behavior: An fNIRS study. *J. Integr. Neurosci.* 14, 207–222. doi: 10.1142/S0219635215500120

Jiang, X., Bian, G. B., and Tian, Z. (2019). Removal of artifacts from EEG signals: A review. *Sensors (Switzerland)* 19, 1–18. doi: 10.3390/s19050987

Jin, X., Chen, C., Zhou, X., and Yang, X. (2019). Stimulating the dorsolateral prefrontal cortex decreases the asset bubble: A tDCS study. *Front. Psychol.* 10:1031. doi: 10.3389/fpsyg.2019.01031

JSME (2009). *Sansuu kyoiku shido yogo jiten (Arithmetic education guidance glossary)*, 4th Edn. Tokyo: Japan Society of Mathematical Education.

Kane, M. J., and Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychon. Bull. Rev.* 9, 637–671. doi: 10.3758/BF03196323

Kurczek, J., Wechsler, E., Ahuja, S., Jensen, U., Cohen, N. J., Tranel, D., et al. (2015). Differential contributions of hippocampus and medial prefrontal cortex to self-projection and self-referential processing. *Neuropsychologia* 73, 116–126. doi: 10.1016/j.neuropsychologia.2015.05.002

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33:159. doi: 10.2307/2529310

Larkin, J. H., and Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cogn. Sci.* 11, 65–100. doi: 10.1111/j.1551-6708.1987.tb00863.x

Lau-Zhu, A., Lau, M. P. H., and McLoughlin, G. (2019). Mobile EEG in research on neurodevelopmental disorders: Opportunities and challenges. *Dev. Cogn. Neurosci.* 36:100635. doi: 10.1016/j.dcn.2019.100635

Lewis, A. B., and Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *J. Educ. Psychol.* 79, 363–371. doi: 10.1037/0022-0663.79.4.363

Manalo, E., and Uesaka, Y. (2016). "Hint, instruction, and practice: the necessary components for promoting spontaneous diagram use in students' written work?," in *Diagrammatic Representation and Inference Diagrams 2018, Lecture Notes in Computer Science*, Vol. 2980, eds M. Jamnik, Y. Uesaka, and S. Elzer Schwartz Cham: Springer, 157–171. doi: 10.1007/978-3-319-42333-3_12

Mayer, R. E., and Gallini, J. K. (1990). When is an illustration worth ten thousand words? *J. Educ. Psychol.* 82, 715–726. doi: 10.1037/0022-0663.82.4.715

Novick, L. R., and Hurley, S. M. (2001). To matrix, network, or hierarchy: That is the question. *Cogn. Psychol.* 42, 158–216. doi: 10.1006/cogp.2000.0746

O'Reilly, R. C. (2010). The What and How of prefrontal cortical organization. *Trends Neurosci.* 33, 355–361. doi: 10.1016/j.tins.2010.05.002

OECD (2018). *The Future of Education and Skills: Education 2030. In OECD Education Working Papers*. Available online at: http://www.oecd.org/education/2030/E2030 Position Paper (accessed April 5, 2018).

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Paas, F., Tuovinen, J. E., Tabbers, H., and Van Gerven, P. W. M. (2011). Cognitive load measurement as a means to advance cognitive load theory cognitive load measurement as a means to advance cognitive load theory. *Educ. Technol.* 38, 37–41. doi: 10.1186/s12909-015-0452-9

Paris, S. G., Lipson, M. Y., and Wixson, K. K. (1983). Becoming a strategic reader. *Contemp. Educ. Psychol.* 8, 293–316. doi: 10.1016/0361-476X(83)90018-8

Paulesu, E., Frith, C. D., and Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature* 362, 342–345. doi: 10.1038/362342a0

Peck, E. M., Yuksel, B. F., Ottley, A., Jacob, R. J. K., and Chang, R. (2013). "Using fNIRS brain sensing to evaluate information visualization interfaces," in *Proceedings of the conference on human factors in computing systems*, (New York, NY), 473–482. doi: 10.1145/2470654.2470723

Peters, L., and De Smedt, B. (2018). Arithmetic in the developing brain: A review of brain imaging studies. *Dev. Cogn. Neurosci.* 30, 265–279. doi: 10.1016/j.dcn.2017.05.002

Petrides, M. (2005). Lateral prefrontal cortex: Architectonic and functional organization. *Philos. Trans. R. Soc. B: Biol. Sci.* 360, 781–795. doi: 10.1098/rstb.2005.1631

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., et al. (2020). The present and future use of functional near-infrared spectroscopy (Fnirs) for cognitive neuroscience. *Ann. N Y Acad. Sci.* 1464, 5–29. doi: 10.1111/nyas.13948

Reis, P. M. R., Hebenstreit, F., Gabsteiger, F., von Tscharner, V., and Lochmann, M. (2014). Methodological aspects of EEG and body dynamics measurements during motion. *Front. Hum. Neurosci.* 8:156. doi: 10.3389/fnhum.2014.00156

Reuter, T., Schnotz, W., and Rasch, R. (2015). Drawings and tables as cognitive tools for solving non-routine word porblems in primary school. *Am. J. Educ. Res.* 3, 1387–1397. doi: 10.12691/education-3-11-7

Riley, M., Greeno, J., and Heller, J. (1983). "Development of children's problem solving ability in arithmetic," in *The Development of Mathematical Thinking*, ed. H. Ginsburg New York, NY: Academic Press, 153–196.

Sakai, J. (2022). Functional near-infrared spectroscopy reveals brain activity on the move. *Proc. Natl. Acad. Sci. U. S. A.* 119, 1–4. doi: 10.1073/pnas.2208729119

Sato, Y., Masuda, S., Someya, Y., Tsujii, T., and Watanabe, S. (2015). "An fMRI analysis of the efficacy of Euler diagrams in logical reasoning," in *Proceedings of the IEEE symposium on visual languages and human-centric computing, VL/HCC, 2015-Decem*, (Atlanta, GA: IEEE), 143–151. doi: 10.1109/VLHCC.2015.7357209

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., and Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instr. Sci.* 43, 93–114. doi: 10.1007/s11251-014-9328-3

Schnotz, W., and Kürschner, C. (2008). External and internal representations in the acquisition and use of knowledge: Visualization effects on mental model construction. *Instr. Sci.* 36, 175–190. doi: 10.1007/s11251-007-9029-2

Sharbrough, F. (1991). American electroencephalographic society guidelines for standard electrode position nomenclature. *J. Clin. Neurophysiol.* 8, 200–202. doi: 10.1097/00004691-199104000-00007

Skau, S., Helenius, O., Sundberg, K., Bunketorp-Käll, L., and Kuhn, H.-G. (2022). Proactive cognitive control, mathematical cognition and functional activity in the frontal and parietal cortex in primary school children: An fNIRS study. *Trends Neurosci. Educ.* 28:100180. doi: 10.1016/j.tine.2022.100180

Smith, E. E., Jonides, J., and Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cereb. Cortex* 6, 11–20. doi: 10.1093/cercor/6.1.11

Soltanlou, M., Sitnikova, M. A., Nuerk, H.-C., and Dresler, T. (2018). Applications of functional near-infrared spectroscopy (fNIRS) in studying cognitive development: The case of mathematics and language. *Front. Psychol.* 9:277. doi: 10.3389/fpsyg.2018.00277

Souma, K. (2020). *Tanoshii Sansuu 6 Nen (Fun Math Grade 6)*, 1st Edn. Tokyo: Dainippon Tosho Publishing.

Strangman, G., Culver, J. P., Thompson, J. H., and Boas, D. A. (2002). A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *NeuroImage* 17, 719–731. doi: 10.1006/nimg.2002.1227

Sugiura, L., Ojima, S., Matsuba-Kurita, H., Dan, I., Tsuzuki, D., Katura, T., et al. (2011). Sound to language: Different cortical processing for first and second languages in elementary school children as revealed by a large-scale study using fNIRS. *Cereb. Cortex* 21, 2374–2393. doi: 10.1093/cercor/bhr023

Suthers, D. D. (2003). Representational guidance for collaborative inquiry. *Arguing Learn* 1994, 27–46. doi: 10.1007/978-94-017-0781-7_2

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* 22, 123–138. doi: 10.1007/s10648-010-9128-5

Sweller, J., Van Merrienboer, J. J. G., and Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205

Uesaka, Y., and Manalo, E. (2006). "Active comparison as a means of promoting the development of abstract conditional knowledge and appropriate choice of diagrams in math word problem solving," in *Diagrammatic Representation and Inference Diagrams 2006, Lecture Notes in Computer Science*, Vol. 4045, eds D. Barker-Plummer, R. Cox, and N. Swoboda Berlin: Springer, 181–195. doi: 10.1007/11783183_25

Uesaka, Y., and Manalo, E. (2012). Task-related factors that influence the spontaneous use of diagrams in math word problems. *Appl. Cogn. Psychol.* 26, 251–260. doi: 10.1002/acp.1816

Uesaka, Y., Manalo, E., and Ichikawa, S. (2007). What kinds of perceptions and daily learning behaviors promote students' use of diagrams in mathematics problem solving? *Learn. Instr.* 17, 322–335. doi: 10.1016/j.learninstruc.2007.02.006

Uesaka, Y., Manalo, E., and Ichikawa, S. (2010). "The effects of perception of efficacy and diagram construction skills on students' spontaneous use of diagrams when solving math word problems," in *Diagrammatic Representation and Inference Diagrams 2010, Lecture Notes in Computer Science*, Vol. 6170, eds A. K. Goel, M.

Jamnik, and N. H. Narayanan Berlin: Springer, 197–211. doi: 10.1007/978-3-642-14600-8_19

Van Garderen, D. (2007). Teaching students with LD to use diagrams to solve mathematical word problems. *J. Learn. Disabil.* 40, 540–553. doi: 10.1177/00222194070400060501

Van Garderen, D., Scheuermann, A., and Jackson, C. (2013). Examining how students with diverse abilities use diagrams to solve mathematics word problems. *Learn. Disabil. Q.* 36, 145–160. doi: 10.1177/0731948712438558

Vogel, S. E., and De Smedt, B. (2021). Developmental brain dynamics of numerical and arithmetic abilities. *NPJ Sci. Learn.* 6, 1–11. doi: 10.1038/s41539-021-00099-3

Wijeakumar, S., Huppert, T. J., Magnotta, V. A., Buss, A. T., and Spencer, J. P. (2017). Validating an image-based fNIRS approach with fMRI and a working memory task. *NeuroImage* 147, 204–218. doi: 10.1016/j.neuroimage.2016.12.007

Xu, J., and Zhong, B. (2018). Review on portable EEG technology in educational research. *Comput. Hum. Behav.* 81, 340–349. doi: 10.1016/j.chb.2017.12.037

Yamaya, N., Tsuchiya, K., Takizawa, I., Shimoda, K., Kitazawa, K., and Tozato, F. (2021). Effect of one-session focused attention meditation on the working memory capacity of meditation novices: A functional near-infrared spectroscopy study. *Brain Behav.* 11, 1–11. doi: 10.1002/brb3.2288

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., and Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nat. Rev. Neurosci.* 20, 364–375. doi: 10.1038/s41583-019-0150-4

Zahner, D., and Corter, J. E. (2010). The process of probability problem solving: Use of external visual representations. *Math. Think. Learn.* 12, 177–204.

Zhang, J. (1997). The nature of external representations in problem solving. *Cogn. Sci.* 21, 179–217. doi: 10.1016/S0364-0213(99)80022-6g an image-based fNIRS appr

Zhang, J., and Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cogn. Sci.* 18, 87–122. doi: 10.1207/s15516709cog1801_3

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 22 November 2022
DOI 10.3389/fpsyg.2022.947768

Check for updates

OPEN ACCESS

EDITED BY
Antonio P. Gutierrez de Blume, Georgia Southern University, United States

REVIEWED BY
Xiaofeng Liu, University of South Carolina, United States
Min Wang, University of Texas at San Antonio, United States

*CORRESPONDENCE
Qianrao Fu
qfu@xauat.edu.cn

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

RECEIVED 19 May 2022
ACCEPTED 04 November 2022
PUBLISHED 22 November 2022

CITATION
Fu Q, Moerbeek M and Hoijtink H (2022) Sample size determination for Bayesian ANOVAs with informative hypotheses.
Front. Psychol. 13:947768.
doi: 10.3389/fpsyg.2022.947768

Sample size determination for Bayesian ANOVAs with informative hypotheses

Qianrao Fu[1]*, Mirjam Moerbeek[2] and Herbert Hoijtink[2]

[1]School of Management, Xi'an University of Architecture and Technology, Xi'an, China, [2]Department of Methodology and Statistics, Utrecht University, Utrecht, Netherlands

Researchers can express their expectations with respect to the group means in an ANOVA model through equality and order constrained hypotheses. This paper introduces the `R` package `SSDbain`, which can be used to calculate the sample size required to evaluate (informative) hypotheses using the Approximate Adjusted Fractional Bayes Factor (AAFBF) for one-way ANOVA models as implemented in the `R` package `bain`. The sample size is determined such that the probability that the Bayes factor is larger than a threshold value is at least η when either of the hypotheses under consideration is true. The Bayesian ANOVA, Bayesian Welch's ANOVA, and Bayesian robust ANOVA are available. Using the `R` package `SSDbain` and/or the tables provided in this paper, researchers in the social and behavioral sciences can easily plan the sample size if they intend to use a Bayesian ANOVA.

KEYWORDS

Bayes factor, Bayesian ANOVAs, informative hypothesis, sample size, SSDbain

1. Introduction

In a classical one-way ANOVA, two hypotheses, the null hypothesis H_0 and the alternative hypotheses H_a are contrasted:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K \quad (1)$$

vs.

$$H_a : \text{not all means are equal}, \quad (2)$$

where μ_k denotes the mean for group $k = 1, 2, ..., K$, and K denotes the number of groups.

Statistical power is the probability to correctly reject the null hypothesis when an effect exists in the population. Cohen (1988, 1992) published some of the most cited literature on power analysis; he proposed the effect size measure $f = \sigma_m/\sigma$, where σ_m denotes the standard deviation of the means of the K groups, and σ the common within-group standard deviation. The classical sample size table of the one-way ANOVA based on the F-test (Cohen, 1992) indicates that in the case of three groups, 322, 52, or 21 subjects per group are needed to obtain a power of 0.8 to detect a small ($f = 0.1$), medium ($f = 0.25$), or large ($f = 0.4$) effect size at a Type I error rate $\alpha = 0.05$.

Required sample sizes for other scenarios can be calculated using software for power analysis and optimal study design, such as `G*Power` (Faul et al., 2007, 2009; Mayr et al., 2007), `nQuery Advisor` (Elashoff, 2007) and `PASS` (Hintze, 2011). Power analysis has become more important in a scientific world with competition for limited funding for research grants. Funding agencies often require value for money: if an effect size exists in the population then it should be detected with sufficient probability. However, many studies in the behavioral and social sciences are underpowered, mainly because of insufficient funding or numbers of subjects willing to participate. As well as a reduced probability of detecting an important effect size, underpowered research causes many problems, including overestimation of effect size, poor replicability of research findings, and thus an increased risk of drawing incorrect conclusions. For relevant articles, see Fraley and Vazire (2014), Maxwell (2004), Simonsohn et al. (2014), Dumas-Mallet et al. (2017), and Szucs and Ioannidis (2017).

Recently, null-hypothesis significance testing (NHST) has been criticized in numerous articles. Unnecessary detail will not be given in this paper, but see the typical references Nickerson (2000), Wagenmakers (2007), Masicampo and Lalande (2012), Harlow et al. (2016), and Wicherts et al. (2016). Alternatives such as Bayesian statistics have as a consequence become increasingly popular over the past decade (Van de Schoot et al., 2017; Vandekerckhove et al., 2018; Wagenmakers et al., 2016). Among them, Bayes factor is the most important tool to evaluate the competing hypotheses. The Bayes factor is the measurement of the relative evidence between two competing hypotheses. For example, if H_0 vs. H_1, and the Baye factor $\text{BF}_{01} = 10$, then the support for H_0 is 10 times more than H_1. The Bayes factor cannot only provide evidence in favor of the alternative hypothesis, but, in contrast to the *p*-value, also provides evidence in favor of the null hypotheses. The Bayes factor quantifies the strength of current data to support for H_0 and H_1, respectively, which is more balanced than the traditional NHST where Bayes factor are more balanced in terms of support for H_0 and H_1, and thus its tendency to reject H_0 is relatively less strong. Under the traditional NHST hypothesis, as long as the collected data is enough the researcher can obtain $p < 0.05$ and thus reject H_0, in contrast to the NHST, the Bayes factor tends to be stable with the increase of data. The Bayes factor does not depend on the unknown or nonexistent sampling plan, while the *p*-value is affected by the sampling plan. In addition, the traditional null and alternative hypotheses as specified by (1) and (2) may not reflect the researcher's expectations. The researcher can express his or her expectations with regard to the ordering of the group means $\mu_1, \mu_2, ..., \mu_K$ in an informative hypothesis (Hoijtink, 2011). For example, consider a comparison of the average body heights of adults in the Netherlands, China, and Japan, as denoted by μ_N, μ_C and μ_J. Informative hypotheses may be formulated on the basis of observations, expectations or findings in the literature. One example is hypothesis $H_1 : \mu_N > \mu_C > \mu_J$. It is worth mentioning that the Bayes factor can not only be used to compare the null hypothesis with alternative hypotheses, but also can be used to compare two informative hypotheses directly. Accordingly, in NHST if ordered hypothesis is included, multiple testing should be carried, which leads to increased chances of false positive results. Software for calculating Bayes factor are the `R` package `BayesFactor`, the `R` package `BFpack`, and the `R` package `bain`, which make the Bayes factor readily accessible to applied researchers. Therefore, it is important that sample size calculations for the Bayesian approach to hypothesis testing become available to researchers in the behavioral and social sciences.

Recently, a sequential Bayesian *t*-test (Schönbrodt et al., 2017) was developed that can, when applicable, avoid an a priori sample size calculation. A sequential test (Wald, 1945) allows researchers to add additional observations at every stage of an experiment depending on whether target strength of evidence is reached. That is, the size of the Bayes factor is large enough or a decision rule whether to i) accept the hypothesis being tested; ii) reject the hypothesis being tested; or iii) continue the experiment by making additional observations is satisfied.

However, a sequential test based on Bayesian updating is not always possible, for example, when the population of research is small (e.g., rare disease or cognitive disorder), when the study is longitudinal and runs for many years, when a research plan with an a priori sample size calculation is to be submitted to an ethical committee, or when researchers want to have an indication of the sample sizes needed even when they do use a sequential design. In these situations sample size determination is necessary. In practice, a combination of sample size determination and Bayesian updating is the best choice. For a more extensive overview of the role of sample size determination and Bayesian updating, the reader is referred to Fu et al. (2020).

Throughout this paper sample size determination (SSD) for the comparison of null, informative, and alternative hypotheses under a one-way ANOVA in the Bayesian framework van den Bergh et al. (2020), which will build on the sample size calculations for *t*-tests discussed in Schönbrodt and Wagenmakers (2018), Stefan et al. (2019), and Fu et al. (2020), will be performed. However, the observed data in social and behavioral research are often non-normal distributed or homogeneous of variance, see, for example, Glass et al. (1972), Micceri (1989), Harwell et al. (1992), Coombs et al. (1996), Keselman et al. (1998), and Blanca et al. (2013). To solve these problems, alternative ANOVAs will also be considered: (1) SSD for Bayesian Welch's ANOVA is available when homogeneity of variance does not hold; (2) SSD for Bayesian robust ANOVA is available when homogeneity of variance and normality of residuals do not hold and/or when the data contain outliers.

The outline of this paper is as follows. First, the models that are used in the article are introduced, the informative hypotheses that are evaluated is described, and the Approximate Adjusted

Fractional Bayes Factor (AAFBF) approach as implemented in the R package bain is elaborated. Subsequently, sample size determination will be introduced, features of SSD will be highlighted, and examples will be provided and discussed. The paper ends with a short conclusion.

2. One-way ANOVAs, (Informative) hypotheses, and Bayes factor

In this paper, K mutually independent group means, $\mu_1, \mu_2, \cdots, \mu_K$ are compared. Three different types of ANOVA models are considered:

Model 1: ANOVA, that is, the within-group variances for the K groups are equal

$$y_{tk} = \sum_{k=1}^{K} \mu_k D_{tk} + \epsilon_{tk}, \epsilon_{tk} \sim N(0, \sigma^2), \quad (3)$$

Model 2: Welch's ANOVA, that is, the within-group variances for the K groups are unequal

$$y_{tk} = \sum_{k=1}^{K} \mu_k D_{tk} + \epsilon_{tk}, \epsilon_{tk} \sim N(0, \sum_{k=1}^{K} \sigma_k^2 D_{tk}), \quad (4)$$

Model 3: Robust ANOVA, that is, the within-group variances for the K groups are unequal, and the distribution of the residuals is non-normal and/or the data contain outliers

$$y_{tk} = \sum_{k=1}^{K} \mu_{k,ROB} D_{tk} + \epsilon_{tk}, \epsilon_{tk} \sim f_k(\epsilon_{tk}), \quad (5)$$

where y_{tk} for person $t = 1, \cdots, N$ belonging to group $k = 1, 2, \cdots, K$ is the dependent variable, N denotes the sample size per group, $D_{tk} = 1$ denotes that person t is a member of group k and 0 otherwise, ϵ_{tk} denotes the error in prediction for person t in group k, $f_k(\epsilon_{tk})$ is an unspecified distribution of the residuals in group k, σ^2 denotes the common within-group variance for each group in case of ANOVA, σ_k^2 denotes the within-group variance of group k in case of the Welch's ANOVA, and $\mu_{k,ROB}$ is the robust estimator of population mean.

In this paper, sample size will be determined under the following situations:

Situation 1: If the researchers believe that nothing is going on or something else is going on but they do not know what, sample size will be determined for the comparison of

$H_0: \mu_1 = \mu_2 = \cdots = \mu_K$ vs. H_a, where H_a: not all means are equal;

Situation 2: Many researchers have clear ideas or expectations with respect to what might be going on. These researchers might believe nothing is going on or have a specific expectation about the ordering of the means. Therefore sample size will be determined for a comparison of

$H_0: \mu_1 = \mu_2 = \cdots = \mu_K$ vs. $H_i: \mu_{1^*} > \mu_{2^*} > \cdots > \mu_{K^*}$;

where $1^*, 2^*, \cdots, K^*$ are a re-ordering of the numbers $1, 2, \cdots, K$;

Situation 3: Or, continuing Situation 2, researchers may want to compare their expectation with its complement. Therefore sample size will be determined for a comparison of

$H_i: \mu_{1^*} > \mu_{2^*} > \cdots > \mu_{K^*}$ vs. H_c: not H_i;

Situation 4: The researchers have two competing expectations

$H_i: \mu_{1^*} > \mu_{2^*} > \cdots > \mu_{K^*}$ vs. $H_j: \mu_{1^\#} > \mu_{2^\#} > \cdots > \mu_{K^\#}$,

where $1^\#, 2^\#, \cdots, K^\#$ denote a re-ordering of numbers $1, 2, \cdots, K$ that is different from H_i. Note that, SSD is also possible if some of the ">" in H_i or H_j are replaced by "=."

The AAFBF as implemented in the R package bain will be used to determine the relative support in the data for a pair of hypotheses. The interested reader is referred to Gu et al. (2018), Hoijtink et al. (2019a) and Hoijtink et al. (2019b) for the complete statistical background. Here only the main features of this approach will be presented. If, for example, $\text{BF}_{ij} = 10$, this implies that the data are 10 times more likely to have been observed under H_i than under H_j. In this manuscript, the AAFBF will be used because it is currently the only Bayes factor available that can handle the four situations introduced above for regular ANOVA, Welch's ANOVA, and robust ANOVA. In what follows, the AAFBF implementation for ANOVAs will be described. First of all, the Bayes factor with which H_0 and H_i can be compared to H_a will be introduced. Subsequently, BF_{ij} and BF_{ic} will be introduced.

Let H_z denote either of H_0 and H_i, and note that for robust ANOVA $\boldsymbol{\mu}$ has to be replaced by $\boldsymbol{\mu}_{ROB}$, then

$$\text{BF}_{za} = \frac{f_z}{c_z} = \frac{\int_{\boldsymbol{\mu} \in H_z} g_a(\boldsymbol{\mu}) d\boldsymbol{\mu}}{\int_{\boldsymbol{\mu} \in H_z} h_a(\boldsymbol{\mu}) d\boldsymbol{\mu}} \quad (6)$$

where f_z and c_z are the fit and complexity of H_z relative to H_a, respectively, $g_a(\boldsymbol{\mu})$ denotes a normal approximation to the posterior distribution of $\boldsymbol{\mu}$ under H_a, and $h_a(\boldsymbol{\mu})$ the corresponding prior distribution of $\boldsymbol{\mu}$ under H_a. The fit is the proportion of the posterior distribution $g_a(\cdot)$ in agreement with H_z, and the complexity is the proportion of the prior distribution $h_a(\cdot)$ in agreement with H_z. The Bayes factor (BF) for H_i against H_j is:

$$\text{BF}_{ij} = \frac{\text{BF}_{ia}}{\text{BF}_{ja}} = \frac{f_i/c_i}{f_j/c_j}, \quad (7)$$

and the BF of H_i vs. H_c is:

$$\text{BF}_{ic} = \frac{\text{BF}_{ia}}{\text{BF}_{ca}} = \frac{f_i/c_i}{(1-f_i)/(1-c_i)}. \quad (8)$$

The posterior distribution used in the AAFBF is a normal approximation of the actual posterior distribution of the K

group means. This can be justified using large sample theory (Gelman et al., 2013, pp. 101). This normal approximation can be specified using the estimates of μ, the residual variance s^2 and N. For the regular ANOVA (Model 1) this renders:

$$g_a(\boldsymbol{\mu}) = \iint_{\mu\in\boldsymbol{\mu}} \pi_a(\boldsymbol{\mu},\boldsymbol{\sigma}^2)\, d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2$$
$$= \int_{\mu\in\boldsymbol{\mu}} \pi_a(\boldsymbol{\mu})\, d\boldsymbol{\mu} = N\left(\left[\ \hat{\boldsymbol{\mu}}\ \right], \begin{bmatrix} \hat{s}^2/N & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \hat{s}^2/N \end{bmatrix}\right); \quad (9)$$

for the Welch's ANOVA (Model 2) this renders:

$$g_a(\boldsymbol{\mu}) = N\left(\left[\ \hat{\boldsymbol{\mu}}\ \right], \begin{bmatrix} \hat{s}_1^2/N & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \hat{s}_K^2/N \end{bmatrix}\right); \quad (10)$$

where $\hat{\boldsymbol{\mu}} = [\hat{\mu}_1, \hat{\mu}_2, \cdots, \hat{\mu}_K]$ denotes the maximum likelihood estimates of the K group means, $\hat{s}^2$ denotes the unbiased estimate of the residual variance, and $\hat{s}_1^2$, $\hat{s}_2^2$, $\cdots$, $\hat{s}_K^2$ denote unbiased estimates of the K within-group variances. For the robust ANOVA (Model 3),

$$g_a(\boldsymbol{\mu}) = N\left(\left[\ \hat{\boldsymbol{\mu}}_{\mathrm{ROB}}\ \right], \begin{bmatrix} \hat{s}_{1,ROB}^2/N & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \hat{s}_{K,ROB}^2/N \end{bmatrix}\right). \quad (11)$$

where $\hat{\boldsymbol{\mu}}_{ROB}$ is the 20% trimmed mean, which according to Wilcox (2017, pp. 45-93) is the best choice, and $\hat{s}_{k,ROB}^2$ is a robust estimate of the residual variance in Group k, which is based on the Winsorized variance (see, Wilcox, 2017, pp. 60-64). If the data are severely non-normal or contain outliers, the estimates of means can be very poor estimates of central tendency, and the within-group variances can be very poor estimates of the variability within a group (Bosman, 2018) therefore in these situations it may be preferable to use $\hat{\boldsymbol{\mu}}_{ROB}$ and $\hat{s}_{k,ROB}^2$ for $k = 1, \cdots, K$.

The prior distribution is based on the adjusted (Mulder, 2014) fractional Bayes factor approach (O'Hagan, 1995). As is elaborated in Gu et al. (2018) and Hoijtink et al. (2019a) for the regular ANOVA with homogeneous within-group variances (Model 1), the prior distribution is:

$$h_a(\boldsymbol{\mu}) = N\left(\left[\ \mathbf{0}\ \right], \begin{bmatrix} \frac{1}{b}\times\frac{\hat{s}^2}{N} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{b}\times\frac{\hat{s}^2}{N} \end{bmatrix}\right); \quad (12)$$

and, for the Welch's ANOVA with group specific variances (Model 2) the prior distribution is

$$h_a(\boldsymbol{\mu}) = N\left(\left[\ \mathbf{0}\ \right], \begin{bmatrix} \frac{1}{b}\times\frac{\hat{s}_1^2}{N} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{b}\times\frac{\hat{s}_K^2}{N} \end{bmatrix}\right); \quad (13)$$

and, for the robust ANOVA (Model 3) the prior distribution is

$$h_a(\boldsymbol{\mu}) = N\left(\left[\ \mathbf{0}\ \right], \begin{bmatrix} \frac{1}{b}\times\frac{\hat{s}_{1,ROB}^2}{N} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{b}\times\frac{\hat{s}_{K,ROB}^2}{N} \end{bmatrix}\right). \quad (14)$$

For the hypotheses considered in this paper mean of the prior distribution should be the origin $\mathbf{0}$. As is elaborated in Mulder (2014), this choice renders a quantification of complexity in accordance with Occam's razor and, as is elaborated in Hoijtink et al. (2019b), it renders a Bayes factor that is consistent. The variances appearing in the prior distribution are based on a fraction of the information in the data. For each group in an ANOVA this fraction is $b = \frac{J}{K} \times \frac{1}{N}$ (Hoijtink et al., 2019a). The choice for the parameter J is inspired by the minimal training sample approach (Berger and Pericchi, 1996, 2004): it is the number of independent constraints used to specify the hypotheses under consideration, because these can be seen as the number of underlying parameters (the differences between pairs of means) that are of interest. Specifically, if $H_0: \mu_1 = \mu_2 = \mu_3$ vs $H_i: \mu_1 > \mu_2 > \mu_3$ is considered, J is equal to 2. The choice for minimum training samples is to some degree arbitrary. It is in general common in Bayesian analyzes to execute sensitivity (to the prior distribution) analyzes. Hence alternative choices of $b = \frac{2J}{K} \times \frac{1}{N}$ and $b = \frac{3J}{K} \times \frac{1}{N}$ are also considered in this paper. Note that, prior sensitivity only applies to Situations 1 and 2, the Bayes factors computed for Situations 3 and 4 are not sensitive to the choice of b (see Mulder, 2014).

3. Sample size determination for one-way ANOVAs

SSD for the Bayesian one-way ANOVA is implemented in the R package `SSDbain`[1]. This section describes the specific ingredients needed for the functions `SSDANOVA` and `SSDANOVA_robust` in the R package `SSDbain`. The interested reader is referred to Appendices A,B for an elaboration of the SSD algorithm. After installing the R package `SSDbain`, the following Call 1 and Call 2 are used to calculate the sample size per group for regular ANOVA and Welch's ANOVA:

Call 1: using Cohen's f (Cohen, 1992) to specify the populations of interest

```
#load SSDbain package
library(SSDbain)
SSDANOVA(hyp1="mu1=mu2=mu3",hyp2="Ha", type="equal",f1
     =0,f2=0.25,var=NULL,
BFthresh=3,eta=0.8,T=10000,seed=10)
```

1 `SSDbain` comes with a user manual and can be installed from https://github.com/Qianrao-Fu/SSDbain. Further information on `bain` can be found at https://informative-hypotheses.sites.uu.nl/software/bain/.

Call 2: using means and variances to specify the populations of interest

```
#load SSDbain package
library(SSDbain)
SSDANOVA(hyp1="mu1=mu2=mu3",hyp2="Ha",type="equal",f1=
      c(0,0,0),f2=
c(5.5,4.5,2),var=c(4,4,4),BFthresh=3,eta=0.8,T=10000,
      seed=10)
```

and the Call 3 below is used for a robust ANOVA:

```
#load SSDbain package
library(SSDbain)
SSDANOVA_robust(hyp1="mu1=mu2=mu3",hyp2="Ha",f1=0,f2
      =0.25,skews=c(0,0,0),
kurts=c(0,0,0),var=c(1.5,0.75,0.75),BFthresh=3,eta
      =0.8,T=10000,seed=10)
```

The following arguments appear in these calls:

1. `hyp1` and `hyp2`, strings that specify the hypotheses of interest. If the unconstrained hypothesis is used, `hyp2="Ha;"` if the complement hypothesis is used, `hyp2="Hc."` In case of three groups the default setting is `hyp1="mu1=mu2=mu3,"` and `hyp2="mu1>mu2>mu3,"` which generalizes seamlessly to more than three groups.
2. `type`, a string that specifies the type of ANOVA. If one expects that the K within-group variances are equal, `type="equal,"` otherwise `type="unequal."`
3. `f1` and `f2`, parameters used to specify the populations corresponding to hyp1 and hyp2, respectively. There are two options. In Call 1 given above `f1` and `f2` denote Cohen's $f = \sigma_\mu/\sigma$ where σ_μ denotes the standard deviation of the means of the K groups, and σ denotes the common within-group standard deviation. If `type = "equal,"` the `var=NULL` is required, where `var = NULL` denotes that the variances do not have to be specified. If `type = "unequal,"` the `var` has to be specified by the users (see the next argument for detail). In Call 2 given above `f1` and `f2` contain the population means corresponding to both hypotheses `hyp1` and `hyp2`. This option can always be used and requires the specification of `var`. In Call 3, the combination of Cohen's f and within-group variances or the combination of means and variances are used to specify the populations of interest. In Appendix C it is elaborated how population means are computed if `f1` and `f2` denote Cohen's f.
4. `var`, vector of length K that specifies the within-group variances of the K groups. If `type = "equal"` and f_1 and f_2 are Cohen's f, the specification `var = NULL` implies that each within-group variance is set to 1. In case of `type = "unequal"` or Call 3, the user needs to input Cohen's f and the variances for each group. The corresponding population means can be computed. In Appendix C it is elaborated how in both cases the corresponding population means are computed.
5. `skews` and `kurts`, vectors of length K that specify the skewness and kurtosis for the K groups compared. Here kurtosis means the true kurtosis minus 3, that is, the kurtosis is 0 when the distribution is normal. The default setting is `skews=c(0,0,0)` and `kurts=c(0,0,0)`, which renders a normal distribution. Note that the relationship kurtosis $\geq$ skewness$^2 - 2$ should hold (Shohat, 1929).

Two situations can be distinguished. If researchers want to execute an ANOVA that is robust against outliers, both `skews` and `kurts` are zero vectors with dimension K. Outliers can be addressed in this manner because robust estimates of the mean and its variance obtained for data sampled from a normal distribution (that is, without outliers) are very similar to the robust estimates obtained for data sample from a normal distribution to which outliers are added. If researchers want to address skewed or heavy tailed data, they have to specify the expected skewness and kurtosis for each group.

The following gives guidelines for choosing appropriate values for skewness and kurtosis. If the population distribution is left-skewed, the skewness is a negative value; if the population distribution is right-skewed, the skewness is a positive value. The commonly used example of a distribution with a positive skewness is the distribution of salary data where many employees earn relatively little, while just a few employees have a high salary. In addition, typical response time data often show positive skewness because long response times are less common (Palmer et al., 2011). The high school GPA of students who apply for college often shows a negative skewness. Furthermore, in psychological research, scores on easy cognitive tasks tend to be negatively skewed because the majority of participants can complete most tasks successfully (Wang et al., 2008). If the population distribution is heavy-tailed relative to a normal distribution, the kurtosis is larger than 0; if the population distribution has lighter tailed than a normal distribution, the kurtosis is smaller than 0.

The values to be used for the skewness and kurtosis can be chosen based on a meta-analysis or literature review (e.g., Schmidt and Hunter, 2015). The absolute value of the skewness is typically smaller than 3 in psychological studies. As a general rule, skewness and kurtosis values that are within ±1 of the normal distribution's skewness of 0 and kurtosis of 0 indicate sufficient normality. Blanca et al. (2013) studied the shape of the distribution used in the real psychology, and found that 20% of the distribution showed extreme non-normality. Therefore, it is essential to consider robust ANOVA when the non-normal distribution is involved. After determining the values of the skewness and kurtosis relevant for their populations, researchers can use `SSDANOVA_robust` to determine the sample sizes needed for a robust evaluation of their hypotheses for data sampled from populations that skewed and/or show kurtosis. The non-normal data is generated from a generalization of the normal distribution that accounts for skewness and kurtosis.

The Tukey g-and-h family of non-normal distributions (see, Headrick et al., 2008; Jorge and Boris, 1984) is commonly used for univariate real data generation in Monte Carlo studies. If the researchers input the skewness and kurtosis, g and h can be obtained (Headrick et al., 2008). The data can be generated as follows. Firstly, T (see point 8 for a explanation on Page 18) data sets with sample size N from the standard distribution are simulated; secondly, observations are transformed into a sample from the g-and-h-distribution as below

if $g \neq 0$

$$T(X) = A + B\exp(h/2X^2)(\exp(gX) - 1)/g \qquad (15)$$

if $g = 0$

$$T(X) = A + B\exp(h/2X^2)X \qquad (16)$$

where $X \sim N(0, 1)$, A is the mean parameter, B is the standard deviation parameter, g is the skewness parameter, and h is the kurtosis parameter.

3.1. Intermezo: The probability that the Bayes factor is larger than a threshold value

In this intermezzo it will be elaborated how the required sample size is determined once the populations corresponding to the two competing hypotheses have been specified, that is, once the population group means, variances, and possibly skewness and kurtosis have been specified. Figure 1 portrays the distributions of the Bayes factor under $H_0: \mu_1 = \mu_2 = \mu_3$ and $H_1: \mu_1 > \mu_2 > \mu_3$, that is, when data are repeatedly sampled from H_0 and for each data set BF_{01} is computed, what is the distribution of BF_{01}, and, when data are repeatedly sampled from H_1 and for each data set BF_{10} is computed, what is the distribution of BF_{10}. Figure 1A shows the distribution obtained using $N = 18$ per group, and Figure 1B shows the distribution obtained using $N = 93$ per group. To determine these sample sizes, two criteria are specified. First of all, what is the required size of the Bayes factor to be denoted by BF_{thresh}; and, secondly, what should be the minimum probability that BF_{01} and BF_{10} are larger than BF_{thresh} denoted by $P(BF_{01} > BF_{thresh}|H_0) \geq \eta$ and $P(BF_{10} > BF_{thresh}|H_1) \geq \eta$, respectively. As can be seen in Figure 1, $BF_{thresh} = 3$ and $\eta = 0.90$, that is, with $N = 18$ $P(BF_{01} > 3|H_0) \geq 0.90$, and with $N = 93$ $P(BF_{10} > 3|H_1) \geq 0.90$. Therefore, to fulfill the criteria for both H_0 and H_1, $N = 93$ persons per group are required.

Two aspects of sample size determination need to be elaborated: how to choose BF_{thresh} and how to choose η. The choice of the BF_{thresh} is subjective, common values are 3, 5, and 10. In high-stakes research, such as a clinical trial to compare a new medication for cancer to a placebo and a standard medication, one would prefer a large BF_{thresh}. In low-stakes research, such as an observational study on the comparison of ages of customers at three different coffeehouses, one may use a smaller BF_{thresh}. The second is how to determine η. It should be noted that 1-η is the Bayesian counterpart of the Type I error rate if `hyp1` is true, and the Bayesian counterpart of the Type II error rate if `hyp2` is true. If the consequences of failing to detect the effect could be serious, such as in toxicity testing, one might want a relatively high η such as 0.90. In studies where one may only be interested in large effects, an error for detecting the effect may not have such serious consequences. Here an $\eta = 0.80$ may be sufficient.

6. `BFthresh`, a numeric value not less than 1 that specifies the required size of the Bayes factor. The default setting is `BFthresh=3`.
7. `eta`, a numeric value that specifies the probability that the Bayes factor is larger than `BFthresh` if either of the competing hypotheses is true. The default setting is `eta=0.80`.
8. `T`, a positive integer that specifies the number of data sets sampled from the populations corresponding to the two hypotheses of interest. A larger number of samples returns a more precise sample size estimate but takes longer to run. We recommend that users start with a smaller number of samples (e.g., `T = 1,000`) to get a rough estimate of sample size before confirming it with the default setting `T = 10,000`.
9. `seed`, a positive integer that specifies the seed of R's random number generator. It should be noted that different data sets are simulated in Step 8 if a different seed is used, and thus, that the results of sample size determination may be slightly different. However, the sample sizes obtained using two different seeds give an indication of the stability of the results (this will be highlighted when discussing Table 4 in Appendix). The default setting is `seed=10`.

The results of the functions `SSDANOVA` and `SSDANOVA_robust` include the sample size required per group and the corresponding probability that the Bayes factor is larger than BF_{thresh} when either of the competing hypotheses is true. For example, if the following call to SSDANOVA is executed

```
library(SSDbain)
SSDANOVA(hyp1 = "mu1=mu2=mu3, "hyp2 = "Ha," type = "
     equal," f1 = 0, f2 = 0.25, var = NULL,
BFthresh = 3, eta=0.8, T = 10000, seed = 10)
```

the results for b based on the minimum value of J, and the results for b based on $2J$ and $3J$ (with the aim to address the sensitivity to the specification of the prior distribution) are:

```
using N = 93 and b = 0.007
P(BF0a>3|H0)=0.977
P(BFa0>3|Ha)=0.801

using N = 83 and b = 0.016
P(BF0a>3|H0)=0.949
P(BFa0>3|Ha)=0.802
```

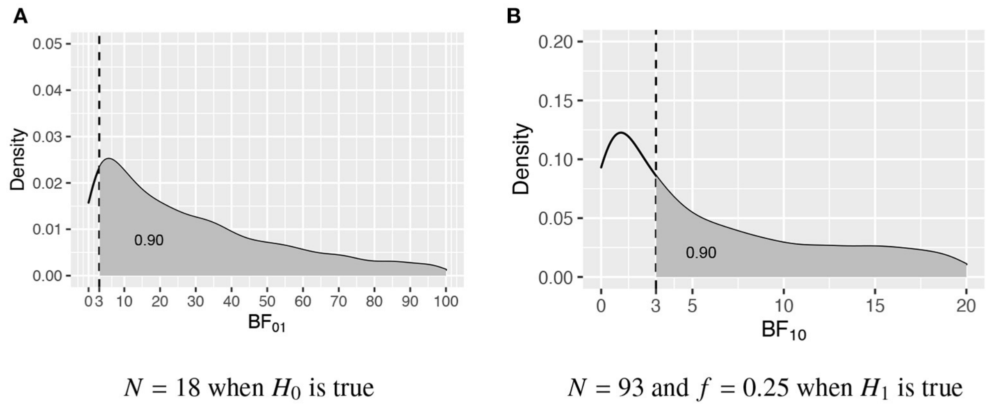


FIGURE 1
The sampling distribution of BF_{01} under H_0 and BF_{10} under H_1. The vertical dashed line represents $BF_{thresh} = 3$, and the gray area denotes η, that is, the probability that the Bayes factor is larger than 3. **(A)** N = 18 when H_0 is true. **(B)** N = 93 and f = 0.25 when H_1 is true.

```
using N = 77 and b = 0.026
P(BF0a>3|H0)=0.918
P(BFa0>3|Ha)=0.802
```

Further interpretation of the results of SSD will be given in the form of three examples that will be presented after the next section.

4. Features of sample size determination for one-way ANOVAs

In this section sample sizes are given based on classical hypotheses, informative hypotheses, and their complement hypotheses for one-way ANOVAs with three groups when the effect size is Cohen's $f = 0.1, f = 0.25$, and $f = 0.4$. Table 1 in Appendix shows the populations corresponding to H_1, H_2, H_a, and H_c for three different effect sizes when the pooled within-group variance is 1. Tables 2–5 in Appendix show the sample size and the corresponding probability that the Bayes factor is larger than BF_{thresh} for regular, Welch's and robust ANOVA for H_0 vs. H_a, H_0 vs. H_1, H_1 vs. H_2, and H_1 vs. H_c, respectively. Table 6 in Appendix displays the robust ANOVA for moderately skewed, extremely skewed, and heavy tailed populations. All the tables are obtained with `set.seed=10`. To illustrate the stability of the results when using `T = 10,000`, in Table 4 in Appendix additionally the results are obtained using `set.seed=1234`. Based on the results presented in these tables a number of features of SSD will be highlighted.

Comparing Table 3 in Appendix with Table 2 in Appendix, it can be seen that the sample size required is smaller if H_0 is compared to the order constrained hypothesis H_1 instead of to the unconstrained hypothesis H_a. For example, if effect size $f = 0.25$, $BF_{thresh} = 3$, $\eta = 0.8$, and regular ANOVA are chosen, the sample size required is 93 per group if H_0 is compared to H_a, while the sample size required is 71 per group if H_0 is compared to H_1. This is because H_1 is more precise than H_a and it is easier to find evidence against or for a more precise hypothesis.

Comparing Table 4 in Appendix with Table 3 in Appendix, it can be clearly seen that the comparison of two non-nested hypotheses like H_1 and H_2 requires less sample size than the comparison of nested hypotheses like H_0 and H_1 (H_0 is in fact on the boundary of H_1). For example, if effect size $f = 0.25$, $BF_{thresh} = 3$, $\eta = 0.8$, and regular ANOVA is used, the sample size required is 71 per group if H_0 is compared to H_1, while the sample size required is 13 per group for H_1 is compared to H_2. The same phenomenon can be observed comparing Table 4 in Appendix (H_1 vs. H_2) with Table 5 in Appendix (H_1 vs. H_c). Although in both cases non-nested hypotheses are compared, H_2 is much more precise than H_c and therefore the required sample size for the comparison of H_1 with H_2 is smaller than for the comparison of H_1 with H_c. In summary the more specific the hypotheses that are evaluated, the smaller the required sample size. The sample size is further reduced if two non-nested hypotheses are compared.

From Tables 2–5 in Appendix, it appears that the sample size required is smaller for a regular ANOVA than for a Welch's ANOVA. For example, as shown in Table 2 in Appendix, if effect size $f = 0.25$, $BF_{thresh} = 3$, $\eta = 0.8$, and H_0 vs. H_a, the sample size required for regular ANOVA is 93 per group, while the sample size required is 102 per group for Welch's ANOVA. However, this is not always the case. The sample size required for Welch's ANOVA may be smaller than the sample size required for a regular ANOVA. The main determinant is order of the size of the variances relative to the order of the means.

For the robust ANOVA, two situations are evaluated. First of all, if the data may include outliers, Tables 2–5 in Appendix apply, because sampling from a normal distribution and using

20% trimming is a very good approximation of sampling from a normal with outliers. Secondly, if the data is skewed or heavy tailed, the results in Table 6 in Appendix apply. Three situations are distinguished: skewness = 0.61 and kurtosis = 0.67, skewness = 1.75 and kurtosis = 5.89, and skewness = 0 and kurtosis = 6.94. These three situations represent moderately skewed, extremely skewed, and extremely heavy-tailed distributions that are often encountered in psychological research (Micceri, 1989; Cain et al., 2017). From Tables 2–5 in Appendix, it can be seen that the sample size required is the largest for robust ANOVA. Comparing Table 3 in Appendix in which the data had a skewness of 0 and a kurtosis of 0 with Table 6 in Appendix, it can be seen that the required sample sizes are larger if robust ANOVA is used to evaluate hypotheses using data sampled from skewed and heavy tailed population distributions.

In addition, the extremely skewed distribution needs smaller sample size than moderately skewed, and the extremely heavy tailed needs a higher sample size than skewed.

Finally, as is illustrated in Table 4 in Appendix, when `T = 1,0000` is used, the results of SSD are very stable, that is, the required sample sizes and η_1 and η_2 are irrelevantly different if different seeds are used. This was also observed for the other tables but these results are not reported in this paper.

5. Examples of sample size determination for one-way ANOVAs

To demonstrate how to use the functions `SSDANOVA` and `SSDANOVA_robust` to execute SSD for one-way ANOVAs in practice, in the following we introduce three practical examples. The first example presents the SSD process for the regular ANOVA, the second example presents the SSD process for the Welch's ANOVA, and the third example presents the SSD process for the robust ANOVA.

`Example 1:` A team of researchers in the field of educational science wants to conduct a study in the area of mathematics education involving different teaching methods to improve standardized math scores. The study will randomly assign fourth grade students who are randomly sampled from a large urban school district to three different teaching methods. The teaching methods are 1) The traditional teaching method where the classroom teacher explains the concepts and assigns homework problems from the textbook; 2) the intensive practice method, in which students fill out additional work sheets both before and after school; 3) the peer assistance learning method, which pairs each fourth grader with a fifth grader who helps them learn the concepts. At the end of the semester all students take the Multiple Math Proficiency Inventory (MMPI). The researchers expect that the traditional teaching group (Group 1) will have the lowest mean score and that the peer assistance group (Group 3) will have the highest mean score. That is,

H_1: $\mu_3 > \mu_2 > \mu_1$.

This hypothesis is compared to H_0 which states that the standardized math scores are the same in the three conditions.

H_0: $\mu_1 = \mu_2 = \mu_3$.

The researchers guess a priori that Group 1 has a mean of 550, Group 2 has a mean of 560, and Group 3 has a mean that equals 580. Based on prior research, the common standard deviation σ is set to 50. Therefore the effect size is $f = \frac{\sigma_m}{\sigma} = 0.249$. The researchers decide to use $\text{BF}_{thresh} = 3$ because they are happy to get some evidence in favor of the best hypothesis. They also choose $\eta = 0.8$ because their research is not a high-stakes research. The researchers also want to do a sensitivity analysis to see how the sample size is influenced by b. To determine the required sample size the researchers use the following call to `SSDANOVA`.

```
library(SSDbain)
SSDANOVA(hyp1="mu1=mu2=mu3," hyp2 = "mu3>mu2>mu1,"
     type = ‘‘equal,’’ f1 = (0,0,0),
f2=c(550,560,580), var = c(2,500,2,500,2,500),
     BFthresh=3,eta=0.8, T = 10000,
seed=10)
```

The results are as follows:

```
using N = 73 and b = 0.009
P(BF03>3|H0)=0.972
P(BF30>3|H3)=0.801

using N = 62 and b = 0.021
P(BF03>3|H0)=0.944
P(BF30>3|H3)=0.803

using N = 55 and b = 0.036
P(BF03>3|H0)=0.909
P(BF30>3|H3)=0.802
```

According to the results the researchers should execute their project using between 55 and 73 persons per group. These are the numbers that they can submit to the (medical) ethical review committee, and, to which they should tailor their resources (time, effort and money). The researchers can combine the results of SSD with Bayesian updating (see the elaboration on this topic in Fu et al., 2020) to avoid using too few or too many persons. Bayesian updating can be executed as follows. They can use 1/4 of the sample size 73, that is, collect 18 students per group firstly, and compute the Bayes factor once the data have been collected. If the Bayes factor is larger than 3, they stop the experiment; otherwise, they collect another 18 students per group, compute the Bayes factor using 36 students per group, and check if the Bayes factor is larger than 3, etc. In this manner, resources can be used in an optimal way while reaching the required amount of evidence.

`Example 2:` A team of psychologists is interested in whether male college students' hair color (1: black, 2: blond, or 3: brunette) influences their social extroversion. The students are given a measure of social extroversion with a range from 0 (low) to 10 (high). Based on a meta analysis of research projects addressing the same research question, the means in the three groups are specified as 7.33, 6.13, and 5.00, and the standard deviations are 2.330, 2.875, and 2.059, respectively. The sampling variance which is denoted as 'var' in the following code is the squared of standard deviation. The effect size is $f = \frac{\sigma_m}{\sigma} = 0.39$. The researchers want to replicate the result emerging of the existing body of evidence, that is, is it H_1: $\mu_1 > \mu_2 > \mu_3$ or H_c: not H_1. They want to obtain decisive evidence $\text{BF}_{thresh} = 10$ with a high probability $\eta = 0.90$. The researchers use the following call to SSDANOVA:

```
library(SSDbain)
SSDANOVA(hyp1="mu1>mu2>mu3," hyp2="Hc," type=''unequal
     ,'' f1=c(7.33,6.13,5.00),
f2=c(5.00,7.33,6.13), var=c(2.330^2,2.875^2,2.059^2),
     BFthresh=10, eta=0.9,
T = 10000, seed=10)
```

The results are as follows:

```
using N = 38 and b = 0.017
P(BF1c>3|H1)=0.903
P(BFc1>3|Hc)=0.988
```

Therefore the researchers should obtain 38 males for each hair color.

Example 3: A team of economists would like to conduct a study to compare the average salary of three age groups in the US. The typical salary distribution in an age group population usually shows positive skewness. Three age groups that include 25-34, 35-44, and 45-54 years old are considered, and the mean salaries for these three groups are denoted as μ_1, μ_2, and μ_3, respectively. Based on prior research, experts' opinion or a pilot study, they assume the effect size is $f = 0.25$, the variances are 1.5, 0.75, and 0.75, the skewnesses are 2, 2.5, and 1.75, and the kurtosis is 6, 10, and 6, respectively. The researchers are only interested in a decision for or against one of the two hypotheses involved. Therefore they use $\text{BF}_{thresh} = 1$ and use $\eta = 0.90$ to have a high probability that the observed Bayes factor correctly identifies the best hypothesis. Two hypotheses are involved: $H_1 : \mu_2 > \mu_3 > \mu_1$ and $H_2 : \mu_3 > \mu_2 > \mu_1$. The following call is used:

```
library(SSDbain)
SSDANOVA_robust(hyp1="mu2>mu3>mu1," hyp2="mu3>mu2>mu1,
     " f1=0.25,f2=0.25,skews=
c(2,2.5,1.75),kurts=c(6,10,6),var=c(1.5,0.75,0.75),
     BFthresh=1,eta=0.9,
T = 10000, seed=10)

using N = 50 and b = 0.013
P(BF23>1|H2)=0.976
P(BF32>1|H3)=0.904
```

The results show that if the researchers survey 50 persons per group, they have a probability that the Bayes factor is larger than 1 of 0.976 if H_1 is true or get a probability that the Bayes factor is larger than 1 of 0.904 if H_2 is true.

6. Conclusion

In this paper we introduced sample size determination for the evaluation of the classical null and alternative hypotheses and informative hypotheses (and their complement) in the one way ANOVA context, using the AAFBF as is implemented in the R package bain. Our SSD approach is implemented in the functions SSDANOVA (which covers regular ANOVA and Welch's ANOVA) and SSDANOVA_robust (which covers robust ANOVA) which are part of the R package SSDbain. Besides the one-way ANOVA, SSDbain also contains the function SSDttest (Fu et al., 2020). In the near future another function, SSDregression, will be added to evaluate (informative) hypotheses using the Bayes factor in the context of multiple regression models. We believe that the R package SSDbain is a welcome addition to the applied researcher's toolbox, and may help the researcher to get an idea about the required sample sizes while planning a research project. The novelty of this research can be concluded as follows:

1. A new sample size determination principle is proposed. Different from traditional unilateral principle, we give a principle, which can be described as the probability that the Bayes factor is larger than a threshold value is at least? when either of the hypotheses under consideration is true.
2. A sample size determination method based on dichotomy is proposed, which can effectively reduce the computation effort. In the traditional sample size determination method, the sample size is increase by 1 until the termination conditions are satisfied. This method is simple and easy to be implemented. However, it might be very time-consuming especially when the sample size is very large. The dichotomy-based sample size determination method only requires a small number of iterations, which is more convenient to the practical application.
3. The sample size determination method proposed in this paper has wider applicability. The software developed in this paper is available for Bayesian ANOVA, Bayesian Welch's ANOVA, and Bayesian robust ANOVA.

The usage of informative hypothesis results in a reduction in the number of sample size required, which further saves the resources. However, Given the sample size requirement for informative hypotheses is usually lower, the researchers may choose to plan their studies with an informative hypothesis even when there is no strong evidence for the specified direction of the means, just so that they can justify their small sample size. This may further exacerbate the replicability crisis problems in the literature. Therefore, the user should be careful if the informative hypothesis is introduced.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

QF, MM, and HH designed the research. QF developed the software package and wrote the paper. MM and HH gave feedback on software development, constructing, and writing the paper. All authors contributed to the article and approved the submitted version.

Funding

QF was supported by the China Scholarship Council. HH was supported by a fellowship from the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS-KNAW) and the Consortium on Individual Development (CID) which was funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO Grant No. 024.001.003).

Conflict of interest

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

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.2022.947768/full#supplementary-material

References

Berger, J. O., and Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *J. Am. Stat. Assoc.* 91, 109–122. doi: 10.1080/01621459.1996.10476668

Berger, J. O., and Pericchi, L. R. (2004). Training samples in objective bayesian model selection. *Ann. Stat.* 32, 841–869. doi: 10.1214/009053604000000229

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., and Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology* 9, 78–84. doi: 10.1027/1614-2241/a000057

Bosman, M. (2018). *Robust Bayes factors for Bayesian ANOVA: Overcoming adverse effects of non-normality and outliers*. Master's thesis.

Cain, M. K., Zhang, Z., and Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: prevalence, influence and estimation. *Behav. Res. Methods* 49, 1716–1735. doi: 10.3758/s13428-016-0814-1

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd Edn.* Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155

Coombs, W. T., Algina, J., and Oltman, D. O. (1996). Univariate and multivariate omnibus hypothesis tests selected to control type i error rates when population variances are not necessarily equal. *Rev. Educ. Res.* 66, 137–179. doi: 10.3102/00346543066002137

Dumas-Mallet, E., Button, K. S., Boraud, T., Gonon, F., and Munafò, M. R. (2017). Low statistical power in biomedical science: a review of three human research domains. *R. Soc. Open Sci.* 4, 160254. doi: 10.1098/rsos.160254

Elashoff, J. (2007). *nQuery version 7.0 Advisor User' Guide.* Los Angeles, CA. Available online at: https://cdn2.hubspot.net/hubfs/488764/nQ70_Manual.pdf

Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fraley, R. C., and Vazire, S. (2014). The n-pact factor: evaluating the quality of empirical journals with respect to sample size and statistical power. *PLoS ONE* 9, e109019. doi: 10.1371/journal.pone.0109019

Fu, Q., Hoijtink, H., and Moerbeek, M. (2020). Sample-size determination for the Bayesian t test and Welch's test using the approximate adjusted fractional Bayes factor. *Behav. Res. Methods* 53, 139–152. doi: 10.31234/osf.io/2v4js

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis, 3rd Edn.* Boca Raton, FL: Chapman and Hall/CRC.

Glass, G. V., Peckham, P. D., and Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Rev. Educ. Res.* 42, 237–288. doi: 10.3102/00346543042003237

Gu, X., Mulder, J., and Hoijtink, H. (2018). Approximated adjusted fractional Bayes factors: a general method for testing informative hypotheses. *Br. J. Math. Stat. Psychol.* 71, 229–261. doi: 10.1111/bmsp.12110

Harlow, L. L. E., Mulaik, S. A., and Steiger, J. H. (2016). *What if There Were no Significance Tests*? New York, NY: Routledge.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., and Olds, C. C. (1992). Summarizing monte carlo results in methodological research: the one-and two-factor fixed effects anova cases. *J. Educ. Stat.* 17, 315–339. doi: 10.3102/10769986017004315

Headrick, T. C., Kowalchuk, R. K., and Sheng, Y. (2008). Parametric probability densities and distribution functions for tukey g-and-h transformations and their use for fitting data. *Appl. Math. Sci.* 2, 449–462. Available online at: https://core.ac.uk/download/pdf/60537699.pdf

Hintze, J. (2011). *PASS 11*. Kaysville, Utah: NCSS, LLC.

Hoijtink, H. (2011). *Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists*. Boca Raton, FL: Chapman and Hall/CRC.

Hoijtink, H., Gu, X., and Mulder, J. (2019a). Bayesian evaluation of informative hypotheses for multiple populations. *Br. J. Math. Stat. Psychol.* 72, 219–243. doi: 10.1111/bmsp.12145

Hoijtink, H., Mulder, J., van Lissa, C., and Gu, X. (2019b). A tutorial on testing hypotheses using the Bayes factor. *Psychol. Methods* 24, 539–556. doi: 10.1037/met0000201

Jorge, M., and Boris, I. (1984). Some properties of the tukey g and h family of distributions. *Commun. Stat. Theory Methods* 13, 353–369. doi: 10.1080/03610928408828687

Keselman, H., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., et al. (1998). Statistical practices of educational researchers: an analysis of their anova, manova, and ancova analyses. *Rev. Educ. Res.* 68, 350–386. doi: 10.3102/00346543068003350

Masicampo, E., and Lalande, D. R. (2012). A peculiar prevalence of p values just below. 05. *Q. J. Exp. Psychol.* 65, 2271–2279. doi: 10.1080/17470218.2012.711335

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychol. Methods* 9, 147–163. doi: 10.1037/1082-989X.9.2.147

Mayr, S., Erdfelder, E., Buchner, A., and Faul, F. (2007). A short tutorial of gpower. *Tutor. Quant. Methods Psychol.* 3, 51–59. doi: 10.20982/tqmp.03.2.p051

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychol. Bull.* 105, 156–166. doi: 10.1037/0033-2909.105.1.156

Mulder, J. (2014). Prior adjusted default Bayes factors for testing (in) equality constrained hypotheses. *Comput. Stat. Data Anal.* 71, 448–463. doi: 10.1016/j.csda.2013.07.017

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychol. Methods* 5, 241–301. doi: 10.1037/1082-989X.5.2.241

O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *J. R. Stat. Soc. B* 57, 99–138. doi: 10.1111/j.2517-6161.1995.tb02017.x

Palmer, E. M., Horowitz, T. S., Torralba, A., and Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *J. Exp. Psychol.* 37, 58–71. doi: 10.1037/a0020747

Schmidt, F. L., and Hunter, J. E. (2015). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. London: Sage.

Schönbrodt, F. D., and Wagenmakers, E.-J. (2018). Bayes factor design analysis: planning for compelling evidence. *Psychon. Bull. Rev.* 25, 128–142. doi: 10.3758/s13423-017-1230-y

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., and Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: efficiently testing mean differences. *Psychol. Methods* 22, 322–339. doi: 10.1037/met0000061

Shohat, J. (1929). Inequalities for moments of frequency functions and for various statistical constants. *Biometrika* 21, 361–375. doi: 10.1093/biomet/21.1-4.361

Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). p-curve and effect size: correcting for publication bias using only significant results. *Perspect. Psychol. Sci.* 9, 666–681. doi: 10.1177/1745691614553988

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., and Wagenmakers, E.-J. (2019). A tutorial on Bayes factor design analysis using an informed prior. *Behav. Res. Methods* 51, 1042–1058. doi: 10.3758/s13428-018-01189-8

Szucs, D., and Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biol.* 15, e2000797. doi: 10.1371/journal.pbio.2000797

Van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., and Depaoli, S. (2017). A systematic review of bayesian articles in psychology: the last 25 years. *Psychol. Methods* 22, 217–239. doi: 10.1037/met0000100

van den Bergh, D., Van Doorn, J., Marsman, M., Draws, T., Van Kesteren, E.-J., Derks, K., et al. (2020). A tutorial on conducting and interpreting a bayesian anova in jasp. *LAnnee Psychol.* 120, 73–96. doi: 10.3917/anpsy1.201.0073

Vandekerckhove, J., Rouder, J. N., and Kruschke, J. K. (2018). Editorial: Bayesian methods for advancing psychological science. *Psychon. Bull. Rev.* 25, 1–4. doi: 10.3758/s13423-018-1443-8

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychon. Bull. Rev.* 14, 779–804. doi: 10.3758/BF03194105

Wagenmakers, E.-J., Morey, R. D., and Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Curr. Dir. Psychol. Sci.* 25, 169–176. doi: 10.1177/0963721416643289

Wald, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat.* 16, 117–186. doi: 10.1214/aoms/1177731118

Wang, L., Zhang, Z., McArdle, J. J., and Salthouse, T. A. (2008). Investigating ceiling effects in longitudinal data analysis. *Multivariate Behav. Res.* 43, 476–496. doi: 10.1080/00273170802285941

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., and Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid p-hacking. *Front. Psychol.* 7, 1832. doi: 10.3389/fpsyg.2016.01832

Wilcox, R. R. (2017). *Introduction to Robust Estimation and Hypothesis Testing, 4th Edn*. New York, NY: Academic Press.

frontiers | Frontiers in Education

TYPE Methods
PUBLISHED 01 December 2022
DOI 10.3389/feduc.2022.965643

Check for updates

OPEN ACCESS

EDITED BY
Christelle Declercq,
Université de Reims Champagne-Ardenne,
France

REVIEWED BY
Jolita Dudaitė,
Mykolas Romeris University,
Lithuania
Salomon Huancahuire-Vega,
Peruvian Union University,
Peru

*CORRESPONDENCE
DingLiang Tan
02088@njnu.edu.cn

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

RECEIVED 10 June 2022
ACCEPTED 14 November 2022
PUBLISHED 01 December 2022

CITATION
Yang X, Dong J and Tan DL (2022) Student conceptual level scale: Development and initial validation.
Front. Educ. 7:965643.
doi: 10.3389/feduc.2022.965643

Student conceptual level scale: Development and initial validation

Xi Yang[1], Jia Dong[2] and DingLiang Tan[2]*

[1]Department of Psychology, Nanjing Normal University, Nanjing, China, [2]Department of Education, Nanjing Normal University, Nanjing, China

The conceptual level is an index of personality development. In the field of teaching, the conceptual level is seen as a dynamic learning style. It has important implications for student learning and individual growth, as well as guidance for teaching. However, the lack of a measurement tool with a clear internal structure for the conceptual level of students has slowed the development of the theory and made it difficult to implement the teaching. To address these issues, this study describes the initial development and validation of the Student Conceptual Level Scale (SCLS) with four samples of students (n=1,321) drawn from eight secondary schools in China. We constructed a second-order three-factor model of the SCLS consisting of three factors—learning awareness level, autonomous input level, and environmental coping level—each with its own independent set of items. This study validated the use of full-scale and subscale scores and examined their relationship with different validity criteria: autonomous learning, mental effort, and academic scores. This updated measure reflects the value and role of the conceptual level in the learning and individual development of students and also provides a more complete frame of reference for the use of the conceptual level in teaching and learning.

KEYWORDS

conceptual level, students, validity, reliability, secondary school

Introduction

The conceptual level (CL) is a personality development index, originally established in the social domain, which refers to an individual's cognitive complexity in response to evolving and changing information (Harvey et al., 1962; Hunt, 1971). CL has been widely used in the field of teaching and learning to represent the learning style of students, which is an indicator of the maturity of students' learning and individual development (Hunt, 1977, 1979; Hunt et al., 1978; Miller, 1981). In general, students perform better across a range of cognitive, emotional, and behavioral domains the higher their CL—for example, showing more complex levels of cognition (Harvey et al., 1962; Hunt, 1971), stronger emotion management skills (Doyle and Rutherford, 1984), more positive attitudes toward learning and more congenial interpersonal relationships (Hunt et al., 1974), less problematic behavior (McLachlan, 1974; Brill, 1978), better environmental adaptability (Hunt, 1977), and increased autonomy (Phillips and Sinclair, 1973). Studies related to the relationship between student CL and learning outcomes have shown that learning is better when students' CL is matched with teacher instruction (Hunt et al., 1974, 1978; Doyle and Rutherford, 1984; Tan, 1995). Matching means that low-CL students are matched with high-structured instruction and high CL students are matched with low-structured instruction. Structure refers to the degree of teacher direction and control over student learning and classroom behavior (Hunt et al., 1974; Miller, 1981).

The concept of CL is derived from conceptual system theory (CST; Harvey et al., 1962; Harold et al., 1967). CST is a theory of personality, and CL is a core element of CST (Hunt, 1975b; Hunt et al., 1978). Influenced by Lewin's (1936) behavioral theory, CST suggests that behavioral outcomes are a function of the interaction between an individual's conceptual level and the environmental conditions he or she is exposed to: When the environmental conditions promote the "cognitive activities" necessary to increase the conceptual level, the individual achieves optimal "growth" (Harvey et al., 1962). Because this viewpoint of CST is based on the personality domain, the CL of personality orientation describes changes in personality development only through the lens of cognitive dispositions and does not focus on behavioral characteristics (Harvey et al., 1962; Miller, 1981). When Hunt extended the ideas of CST to the field of teaching and learning (Hunt, 1970), the CL of learning style orientation described changes in student learning and individual development beyond learning in terms of both cognitive dispositions and behavioral characteristics. CL reflects not only the cognitive complexity of students' perceptions of learning information and information in general but also cognitive projections to behavioral characteristics such as independence in processing information and flexibility in responding to complex situations (Hunt, 1977; Hunt et al., 1978; Miller, 1981). In fact, regardless of the orientation, cognitive and behavioral characteristics cannot be separated diametrically when discussing individual characteristics. Hunt (1975b, 1977) described CL in terms of cognitive complexity, independence, and environmental adaptability, which relate to both cognitive dispositions and behavioral characteristics. Cognitive complexity refers to the ability to process information and reflects the extent to which students' cognition develops on the concrete–abstract dimension; independency refers to the ability to learn independently and reflects the extent to which students develop on the dependence–independence dimension; environmental adaptability refers to the adaptation to the learning environment and reflects the extent to which students encounter difficult or contradictory situations and the degree of development on the conflict–balance dimension (Hunt, 1977; Miller, 1981). From these three perspectives, high-CL students are characterized by a high level of cognitive complexity and the ability to solve more abstract problems; the ability to think and learn independently, with less reliance on outside help; and a high level of environmental adaptability, which allows them to better achieve a balance between internal and external environments when they encounter frustration or conflict. Students with low CL are characterized by one-sided cognition, dependence on the outside world, and difficulty balancing their relationship with the outside world; they need more help from the outside world to learn and grow more effectively (Hunt et al., 1978; Miller, 1981). CL theory suggests that instruction should be matched to the student's CL to facilitate further improvement in learning and CL and to move closer to the developmental goal of "autonomy" (Hunt et al., 1978), which is especially important for students at low conceptual levels (Tomlinson and Hunt, 1971; McLachlan and Hunt, 1973).

CL has potential educational value: (1) It has universal significance for individual student development—it specifies the maturity level of students at this stage of learning and individual development. (2) It provides the basis for the implementation of teaching—teaching based on students' conceptual level can better facilitate students' learning and personal development and promote students' growth while optimizing the teaching and learning process and preventing under-resourcing or waste (Brophy and Good, 1974; Hunt, et al., 1974; Hunt, 1977). Hunt (1975c, 1976b), even found that CL had an impact on teachers' teaching levels: Students at high conceptual levels had a "student pull" on teachers, which allowed teachers to unconsciously adapt their teaching to students' needs, and teachers increased their "adaptive level."

For the measurement of CL, there are currently only three measurement instruments, all of which are projective tests: TIB (This I Believe; Harvey et al., 1962; Greaves, 1971; Stoppard and Miller, 1985), ITI (Interpersonal Topical Inventory; Tuckman, 1965), and, adapted from the PCT (Paragraph Completion Test), PCM (Paragraph Completion Method; Hunt et al., 1978). Among them, the TIB and ITI, both of which are biased by their scale structure and theoretical content, have low correlations, and they are not used to measure conceptual level as a student learning style (Miller, 1978). For example, the TIB is used to measure the teacher's CL (Murphy and Brown, 1970) or the CL of individuals outside of the teaching context (Goldberg, 1974). The topics, formats, and scoring criteria are the same for both PCM and PCT, except for the scoring range: 1–7 for PCT and 0–3 for PCM (Miller, 1981). The only widely used measurement tool for student CL is the PCM. The PCM is a semi-projective measure that consists of six themes, with each item giving only the first part of the question and requiring the respondent to complete the rest of the question. These themes are designed to understand how students respond to conflict or ambivalent situations (e.g., "criticized," "unsure," or "disagree") and their perceptions of rules and authority complexity ("rules," "parents," "being told"). The PCM has a number of limitations: (1) In terms of instrument characteristics, the scoring results of projective tests are difficult to objectify, the administration procedures are complex and inefficient, and they are suitable for case studies (Dai et al., 1999, pp. 241, 242). (2) The PCM was originally used to assess personality, and when used to assess learning styles, it was not possible to demonstrate the generalizability of the instrument across contexts, especially on the "authority, rules" theme (Miller, 1981). (3) The reliability and validity are not satisfactory (Gardiner, and Schroder, 1972; Miller, 1978). (4) Different researchers disagree in the analysis of the data. For example, when grouping scores high and low, some scholars have used the median as a basis for classifying high and low levels of CL, showing that the mean score of the high-CL group was instead lower than the mean score of the low-CL group. Such is a false negative result (Miller, 1981). Other researchers refer to both the range of the score distribution (large enough) and the absolute level (Hunt et al., 1974; McLachlan, 1974). Therefore, as an unstructured measurement tool, it is difficult for the PCM reflect the internal structure of CL. This inconsistency between the theoretical connotation and the measurement tool has created a situation where the conceptual level is difficult to generalize. This is also a great challenge for CL to exert its value. It can be seen that constructing a measurement tool with good measurement properties is the foundation of subsequent CL research.

The present study was based on the theoretical connotations of CL as a student learning style (Hunt et al., 1974, 1978; Hunt, 1975a) to develop a measurement instrument and used secondary school students as the study population. This was for the following reasons. (1) From the perspective of cognitive development theory (Piaget, 1953), secondary school students are in the formal operational stage of cognitive development. This stage reflects the transformation from concrete to abstract thinking: secondary school students process information cognitively in a more profound and abstract way. (2) According to Erikson's (1959) self-concept theory, secondary school students are in a period of self-identity formation and identity confusion and are acquiring self-identity, during which individuals face more internal confusion and conflict and struggle to find a sense of self-continuity and coherence. During this period, adolescents struggle to construct a balance between the self and the external environment (Nie and Ding, 2009). In this way, the cognitive development of secondary school students is markedly complex and conflictual, and they vary more markedly in three aspects of CL (conceptual complexity, independence, and environmental adaptation; Hunt, 1975a; Hunt et al., 1978). Therefore, the development of a measurement instrument for secondary school students' CL has both theoretical and practical implications: Theoretically, it helps to clarify the internal structure of students' CL; practically, it helps to gain insight into students' cognitive and behavioral development and thus provide targeted educational responses.

Based on the connotation of learning style of CL, this study proposed the concept and dimensions of student CL through qualitative research around the characteristics of three aspects of CL (conceptual complexity, independence, and environmental adaptability), developed the Student Conceptual Level Scale (SCLS), and conducted a large-sample reliability test on the scale to confirm the necessity and value of distinguishing CL into more factors.

Study 1

The aim of Study 1 was to generate a pool of items that comprehensively structure and measure the dimensions of SCLS through qualitative research and to examine the content of the items.

Methods

Participants

Personal interviews and open-ended questionnaires were used for initial project construction. A total of seven secondary school students in three provinces and cities were interviewed personally (five from Shanghai, one from Inner Mongolia, and one from Hunan; aged 12–14 years). An open-ended questionnaire was administered to 32 secondary school students in 3 provinces and cities (10 in Shandong, 20 in Ningxia, and 2 in Hainan; ages 12–15 years, grades 7–9, 32 valid questionnaires). All studies reported in this article were approved by the ethics committee.

The expert panel was recruited to assess the content and face validity of the generated items. The expert panel consisted of 10 experts, including 2 professors of psychology and 8 PhDs in psychology.

The pretest and recheck items were designed to determine the official initial version of the scale for use in the exploratory factor analysis (EFA). The pre-test consisted of 20 secondary school students recruited online (age 12–15 years, grades 7–9, 20 valid questionnaires). Re-checking of the initial project consisted of six experts (two PhDs in Psychology, two graduate students in Chinese, and two secondary school psychology teachers).

Procedure

The qualitative study used a combination of personal interviews and open-ended questionnaires. (a) Based on the results of personal interviews and open-ended questionnaires, the detailed records of interview and open-ended questionnaire results were compiled, their contents were analyzed, and representative terms were extracted for classification and summarization. (b) The themes were further refined and compared and analyzed with literature collation, items with high overlap and relevance were retained, the empirical structure of CL was established, and self-edited question items were generated. (c) By extracting scales related to CL learning connotations, similar meaningful questions were adapted to generate adapted items. (d) After generating an initial pool of questions consisting of both self-edited items and adapted items, the content and face validity of the items were examined by a panel of experts. Ten experts iteratively revised all items, modifying the expression of some entry statements to ensure that there were no biases in the subjects' understanding of the questions. (e) To ensure that the question items could be accurately understood by the respondents without bias, 20 secondary school students were recruited through the Internet to make small-scale predictions. (f) After the end of the prediction, six experts were asked to revise the content and language of the questionnaire for items with quality issues (e.g., not easily understood, ambiguously expressed, and sensitive) and then separately asked to review the clarity and appropriateness of the revised questions.

Measures

Interview outline: (1) How do you understand the concept of learning? (2) What are your expectations of learning? (3) How do you think about the rules of school? (4) How do you view yourself in relation to others? (5) How do you face the temptations in learning? (6) How do you feel about what others say about you? (7) Which teaching style is conducive to improving your academic performance? (8) Please give a comprehensive evaluation of yourself.

Open-ended questionnaire: (1) What do you consider to be independent, and please summarize it in one or two sentences. (2) How do you understand "independence in learning"? In what aspects of learning are you independent? In what ways are you dependent on others (teachers, parents)? (3) What factors might affect your independence in learning? (4) Which of the following two teaching styles do you think is more helpful to you in your learning: (a) a teaching style in which the teacher talks a lot and strictly manages the classes, or (b) a teaching style in which the teacher talks less and is loosely managed? What do you think are the benefits of the teaching style you have chosen? (5) How would you assess your own performance in learning? If you could change your learning, what would you like to change (e.g., try to control negative emotions, learn more consciously, try to understand different perspectives, be brave enough to ask questions, etc.)? (6) Do you ignore rules (including school rules and regulations, classroom discipline) when they differ from your personal views? Why? (7) How do you think good or bad relationships with others (parents, teachers, classmates) affect your learning? Please give your opinion based on this. (8) Do you study hard? Why? (9) Please combine the above questions and give a comprehensive evaluation of your value, and please talk about your ideas in as much detail as possible.

Analysis and results

Based on the connotation of CL learning style, all contents related to the three aspects of CL (cognitive complexity, independence, and environmental adaptability) were included in the individual interviews to avoid bias caused by the researcher's subjective screening. In the open-ended questionnaire, the nine questions were organized around six areas: evaluation of one's own independence; factors influencing learning independence; evaluation of the choice of teaching methods; evaluation of one's own learning; evaluation of learning environment factors (others, institutions, interpersonal relationships); and learning motivation and self-evaluation. The results of the qualitative analysis showed that the structure of the student CL involves a total of three aspects: learning awareness (cognition), autonomous input (independence), and environmental coping (environmental adaptation). Therefore, this study constructed the initial dimensions of CL from three aspects—the level of learning awareness, the level of autonomous input, and the level of environmental coping—and generated 60 self-edited question items based on these 3 aspects. The adapted items were derived from the Middle School Students' Self-control Ability Questionnaire (Wang and Lu, 2004), the Personal Goal Orientation Questionnaire (Li and Lin, 2001), the Influences of Middle School Students' Informal Reasoning Questionnaire (Zhang et al., 2010), and the Adolescent Self-consciousness Scale (Nie and Ding, 2009). Six question items were eventually generated. After 66 initial question items were generated, reviewed, and revised by 10 experts, and after an overall agreement of 90%, the SCLS (initial version) was developed for this study. The SCLS (initial version) resulted in an initial scale of 51 questions corresponding to 3 dimensions: 18 questions for the learning awareness dimension, 14 questions for the autonomous input dimension, and 19 questions for the environmental coping dimension. Twelve of the questions were reverse scored. A six-point Likert scale was used, ranging from "not at all" to "completely." The total score of the CL is the sum of the scores of each dimension. The higher the total score, the higher the CL.

Study 2

The aims of Study 2 were to explore the factor structure of the SCLS (initial version), including item analysis and EFA, and to assess the factor structure of the SCLS retained from EFA and compared with other models and validity scales, including confirmatory factor analysis (CFA) and scale reliability tests.

Methods

Participants

The questionnaire was administered to students in eight secondary schools in six provinces in China, and the subjects did not include any from the pretest stage. Questionnaires were considered invalid if: *a*. They included proactive approaches (Dunn et al., 2016) containing self-reported items, such as "I did not answer the above questions seriously" and "I answered the above questions truthfully according to my situation"; if these were answered incorrectly, the responses were deleted (Huang et al., 2012). *b*. More than half of the answers were blank. *c*. They had straight-lining (Curran, 2015; Fang et al., 2016), including both the choice of the same number throughout (e.g., 111,111) or a snake arrangement (e.g., 123,321). Cluster sampling and stratified sampling were used to obtain 4 samples, 1,553 questionnaires were distributed, invalid questionnaires were deleted, and finally, 1,321 valid questionnaires were obtained, with an effective rate of 85%. The specific sample composition is shown in Table 1.

Sample 1: A total of 624 secondary school students completed the SCLS (initial version), which was used for EFA. Among them, students in one secondary school in Ningxia Province received a web-based survey, while the rest of the students received an on-site questionnaire.

Sample 2: A total of 385 secondary school students completed the SCLS (formal version) for CFA. All students were administered the on-site questionnaire. Of these students, 150 were additionally required to fill out the Cognitive Load Questionnaire (CLQ), which was used to measure the scale validity.

Sample 3: A total of 226 secondary school students completed the SCLS (formal version) and the Motivated Strategies for Learning Questionnaire (MLSQ, Chinese version) for measuring the scale validity.

Sample 4: A total of 86 secondary school students completed the SCLS (formal version) for retest reliability analysis, all of whom received on-site questionnaires.

The students were recruited in a class setting, with the prior authorization of the teacher. Once written informed consent was obtained from every student, the survey was given in pen and paper format or *via* online questionnaire on a computer at the end of class.

TABLE 1 Sample characteristics.

Sample	1	2	3	4	Total
Recruited *n*	724	466	268	95	1,553
Valid *n*	624	385	226	86	1,321
Age *M* (*SD*)	16.75	14.68	14.45	14.80	15.62
	(*5.86*)	(*0.76*)	(*0.94*)	(*0.36*)	(*4.20*)
Gender (% *M*)	45.83	52.21	54.42	56.98	49.89
Area (%)					
Gansu	14.90				
Henan	50.96				
Ningxia	34.13				
Inner Mongolia		23.12			
Hainan		76.88		100.00	
Jiangsu			100.00		
Secondary School Grade (%) (Grade)					
7	16.99	57.40	34.07		30.58
8		21.30	65.93	100.00	24.00
9	33.97	21.30			22.26
10	34.13				16.12
11	14.90				7.04

Valid *n* = participants who passed screening criteria and completed the study. Percentages by school grade group reflect the proportion of valid *n* in the respective sample. Grade 7 is the first year of secondary school.

Measures

Cognitive load

The CLQ uses a six-point Likert scale ranging from 1 to 6 for 15 self-reported items. The CLQ includes three dimensions, i.e., mental engagement, emotional engagement, and time engagement. The three dimensions measure "mental effort" in the cognitive load (Zhao, 2011). Mental effort is the core component of cognitive load and reflects effort and engagement in learning (Gerjets et al., 2009; Leppink et al., 2013; Chang et al., 2017). CL also reflects students' engagement in and management of learning (Hunt, 1976a, 1977; Miller, 1981). Therefore, the present study hypothesized that CL is related to mental effort. In previous studies, the CLQ had good reliability and validity indicators (Li and Luo, 2014; Sun and Liu, 2016; Sun, 2016), as well as good internal consistency reliability and validity of the validity scales (Cronbach's α of 0.798 for the total scale and 0.614 to 0.721 for the sub-scale), with good construct validity. In this study, Cronbach's α of the CLQ was 0.87, and that of mental engagement, emotional engagement, and time engagement was 0.79, 0.81, and 0.84, respectively.

MSLQ (Chinese version)

Two subscales of the MSLQ (Chinese version) that reflect autonomous learning abilities, Strategy Use (13 items) and Self-Regulation (9 items), were selected for this study, with a total of 22 questions and 4 reverse-scored items. The MSLQ (Chinese version) uses a seven-point Likert scale ranging from 1 to 7 (Rao and Sachs, 1999). Both autonomous learning and CL can reflect the extent to which students are self-responsible in their learning (Hunt et al., 1974; Wirth et al., 2020), and both include core characteristics such as independence and self-awareness (Hunt, 1976a; Wong et al., 2018; Carpenter et al., 2020). Therefore, this study hypothesized that CL is associated with autonomous learning. In previous studies, the MSLQ (Chinese version) had good reliability and validity indicators (Cronbach's α of 0.92; Li, and Yin, 2010; Lin et al., 2013) and also had good construct validity. In this study, the MSLQ (Chinese version) had a Cronbach's α of 0.89, and Strategy Use and the Self-Regulation had Cronbach's α of 0.82 and 0.80, respectively.

Academic scores

Academic scores refer to the final exam results of the semester, including three subjects: Chinese, math, and English. CL is an individual characteristic that, in general, has a direct relationship with academic scores (Ghazivakili et al., 2014; Yazici, 2016; İlçin et al., 2018). Therefore, this study hypothesized that CL is related to academic scores.

Data analysis

Using SPSS 25.0 software, item analysis, correlation analysis, and EFA were performed for the data from sample 1, and reliability analysis and CFA were performed for the data from sample 2. First, item–total statistics were used to test whether all the items were consistent with the scale. Inconsistent items were removed based on the results. Second, the Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity were used to test whether the data were appropriate for factor analysis. Third, the latent structure of the SCLS item set was explored by combining three principal factor extraction methods: principal component analysis (PCA), varimax, and parallel analysis (PA). The criteria for dimensions and item reduction were as follows (Fan et al., 2003; Wu, 2018): (1) eigenvalues >1; (2) factors containing three or more items; (3) items' loading onto factors was sufficiently large (>0.40); and (4) items did not cross-load onto other factors (other factors <0.35). For the sample 2 data, Mplus 8.0 software was used for CFA.

Results

EFA

Item analysis

The aim of item analysis is to test the reliability of the items in the scale. The critical ratio (CR) of the sample 1 data and the item–total score correlation coefficients were used for item analysis of the 51 initial items using correlation analysis. First, the CR of the extreme groups was calculated, and the total scores were grouped into high and low groups. The first 27% of the total scores were considered the high group, the last 27% of the total scores were considered the low group, and the significance test of differences was done for the total and mean scores of the two extreme groups. Combining the two criteria with a significance level below 0.05 and CR <3.00, the four question items that did not meet the criteria were removed. Second, 11 question items were removed according to 2 criteria of correlation coefficients >0.3 and significance at the 0.05 level between each item and the total score of the scale, after which the test was conducted again.

Results of item analysis

The results showed that the correlation coefficients of each item with the total score ranged from 0.30 to 0.64. The skewness of all items ranged from −0.953 to 0.426, which did not exceed the critical value of 3. The kurtosis was −1.287 to 0.829, which was not >10, indicating that the scores of all questions conformed to a normal distribution. Meanwhile, the skewness and kurtosis were within the acceptable value range of ±2 (Trochim and Donnelly, 2008; Gravetter and Wallnau, 2014), indicating that none of the items had a substantial ceiling or floor effect.

Factor analysis

Based on the results of item analysis, EFA was performed on the 36 retained items. To ensure the feasibility of EFA, Bartlett's sphericity test and the KMO test were performed before each EFA was done. The results of the first EFA showed that the KMO = 0.93 and Bartlett's spherical test chi-square value ($\chi^2_{(630)} = 6868.03$, $p < 0.001$) reached a significant level, indicating that the sample data were suitable for EFA (Hutcheson and Sofroniou, 1999; Field, 2013). PCA and varimax were selected to perform the extraction of common factors (Figure 1) according to the principle of an eigenvalue >1 (Fan et al., 2003), and combined with the gravel plot and PA test (O'connor, 2000; Hayton et al., 2004). Finally, 13 valid items were generated, and a total of 3 factors with eigenvalues >1 were obtained, explaining a total of 51.46% of the total variance. The PA results showed (Figure 1) that the first three factors in the actual data had greater eigenvalues

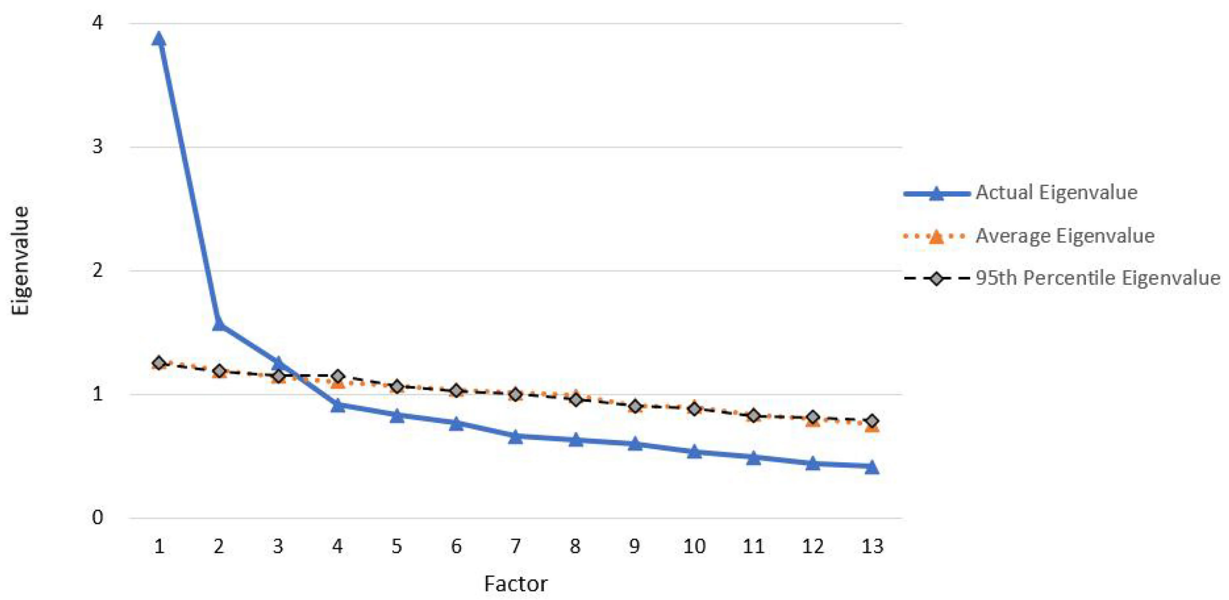


FIGURE 1
Parallel analysis of SCLS. SCLS, Scale of Student Conceptual Level.

than the average eigenvalue and 95th percentile eigenvalue of the random matrix and had a larger retention value. Meanwhile, the remaining factors were smaller than the average eigenvalue and 95th percentile eigenvalue of the random matrix, with relatively small retention value. The results of the two methods were consistent.

The final determination of the three extracted factors was consistent with theoretical expectations, reflecting three aspects of the CL: cognitive complexity, independence, and environmental adaptability. The scale factor loadings are shown in Table 2. According to the content reflected by the items contained in each factor, factor 1 was named learning awareness, which refers to the degree of conscious awareness of learning goals, plans, and tasks. It reflects the individual's level of awareness of learning itself and his or her own learning state and contains five items. Factor 2 was named autonomous input, which refers to the individual's mental ability to consciously inhibit distractions, resist temptations, and think deeply during learning, reflecting the level of independent learning. Containing four items, factor 3 was named environmental coping, which refers to the level of dealing with conflict situations and difficult situations and reflects the degree of individual adaptation to the environment. Based on the connotation of the three factors, this study defines CL as the level of learners' psychological development in three areas: learning awareness, learning behavior, and processing information from the internal and external environment.

TABLE 2 Exploratory Factor Analysis (EFA) on Final 13 Student Conceptual Level Items (*n*=624).

(Sample 1)

Items	**Factor**		
	Awareness[a]	**Input[b]**	**Coping[c]**
I'm unlikely to give up my study goals easily	0.73	0.12	0.26
I always pay attention to the teacher's requirements in class	0.71	0.10	0.06
I take notes in class even if the teacher does not require it	0.71	0.31	(0.07)
I think about how to plan my study and make specific steps accordingly	0.68	0.13	0.18
I know exactly what my study goal is for this semester	0.66	0.10	0.33
It's hard to start studying when I'm addicted to TV or games	0.11	0.78	0.07
I have a hard time getting into a study state when I get excited in play	0.14	0.77	0.08
My mind often goes off when I'm doing homework	0.12	0.67	0.13
I rarely thinking further when I've understood what the teacher said	0.14	0.59	0.02
I do not get angry even if someone disagree with my views	0.01	0.14	0.75
I will not lose heart even if the exam result is lower than my expectations	0.07	0.02	0.66
In a discussion, I will listen to others patiently even if I think they are wrong	0.22	0.12	0.65
I do not interrupt someone who is expressing wrong opinions	0.29	0.00	0.48

Factor loadings in bold (>0.4) represent items loading onto a specific factor. SCLS, Scale of Student Conceptual Level. SCLS accounts for 51.46% of the variance.[a]Factor accounts for 29.69% of variance.
[b]Factor accounts for 12.04% of variance.
[c]Factor accounts for 9.74% of variance.

CFA

To verify the structural validity of our scale, the structural equation modeling statistical software Mplus 8.0 was used for CFA on the sample 2 data. To evaluate the CFA models, we considered four fit statistics as primary indicators of model fit: the Tucker Lewis index (TLI; Tucker and Lewis, 1973), the comparative fit index (CFI; Bentler, 1990), the standardized root mean square residual (SRMR; Pavlov et al., 2021), and the root mean square error of approximation (RMSEA; Browne and Cudeck, 1992) due to the sensitivity of the chi-square statistic to sample size.

Results of CFA

The results of CFA are shown in Table 3. Among the three factors, the fitted indicators for each of the learning awareness factors were $\chi^2=10.192$, $df=5$, CFI = 0.979, TLI = 0.958, RMSEA = 0.052, and SRMR = 0.031. The fitted indicators for each of the autonomous input factors were $\chi^2=13.443$, $df=2$, CFI = 0.990, TLI = 0.970, RMSEA = 0.043, and SRMR = 0.020. The fitted indicators for each of the environmental coping factors were $\chi^2=32.033$, $df=6$, CFI = 0.798, TLI = 0.663, RMSEA = 0.10, and SRMR = 0.089. The scale was validated, and the first-order three-factor model underperformed on all indicators: $\chi^2=253.582$, $df=65$, CFI = 0.721, TLI = 0.665, RMSEA = 0.087, SRMR = 0.075. The second-order three-factor model fit was excellent, and the indicators were $\chi^2=100.461$, $df=62$, CFI = 0.943, TLI = 0.928, RMSEA = 0.040, and SRMR = 0.046. The data statistics showed that the second-order three-factor model fit better than the first-order three-factor model and was more consistent with theoretical expectations, indicating that the SCLS had good structural validity, and the later analysis was based on this model.

TABLE 3 SCLS of Fit of CFAs for Sample 2 (*n*=385).

Model	χ^2	***df***	**TLI**	**CFI**	**SRMR**	**RMSEA (90% CI)**
Correlated 3-factor	253.582	65	0.665	0.721	0.075	0.087 [0.076, 098]
2-order 3-factor	100.461	62	0.928	0.943	0.046	0.040 [0.025,0.054]

CFA, confirmatory factor analysis; CFI, comparative fit index; TLI, Tucker–Lewis index; SRMR, standardized root mean square residual; RMSEA, root mean square error of approximation.

TABLE 4 Partial Correlations, Cronbach's Alpha, Retest Reliability, and CR (and AVE' Sqr for) Between SCLS' Full-Scale and Subscale Scores.

Factors	Cronbach's α	Retest reliability	Spearman–Brown	CR	1	2	3
1. Awareness	0.73	0.81	0.75	0.83	*(0.70)*		
2. Input	0.65	0.67	0.61	0.80	0.33***	*(0.71)*	
3. Coping	0.60	0.67	0.60	0.64	0.40***	0.31***	*(0.70)*
Total CL	0.77	0.79	0.68	0.92	0.79***	0.73***	0.73***

Test–retest reliability ($n = 86$), Cronbach's alpha, Spearman–Brown, and CR ($n = 385$) of each factor; CR, composite reliability. AVE' Sqr = square root of the average variance extracted (italicized in parentheses). Partial correlations after removing variances of gender, age, and grade are shown in italics. Awareness = learning awareness; input = autonomous input; coping = environmental coping.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Reliability and validity assessment

Partial correlations and reliability test

First, sample 2 ($n = 385$) was analyzed for partial correlation between the total scale and 3 subscales. Second, Cronbach's α, retest reliability, Spearman–Brown, and CR were tested for the total scale and three subscales. In addition, 95 subjects in sample 4 were randomly selected (valid $n = 86$), and the interval between the pre-and post-tests was 3 weeks to test the retest reliability of SCLS. The main reliability indexes are shown in Table 4. All the reliability indexes of SCLS met the acceptable standard of the reliability coefficient (Wu, 2018, P244). The square root of the average variance extracted (AVE) of the three subscales was greater than the correlation coefficients of 0.33, 0.40, and 0.31, indicating good discriminant validity. In conclusion, when evaluating the CL of students, it is better to refer to the SCLS total score and the scores of the three high-order factors, rather than explain the scores of the 13 low-order factors alone.

Criterion validity

Since the method used to collect data in this study was self-reporting, there was a possibility of common method variance (CMV). Therefore, the Harman single-factor test was used for both sample data. The results showed that there were 11 factors with eigenvalues >1 in sample 2, and the variance explained by the first factor was 17.43%, which was less than the critical criterion of 40%. In sample 3, there were eight factors with eigenvalues >1, and the first factor explained 28.94% of the variance. Both samples were <40% of the critical threshold, indicating that there was no serious CMV in this study (Zhou and Long, 2004). In the absence of severe CMV, the study then tested the correlations of the three factors of the SCLS with Chinese, math, and English scores, respectively, and the results are shown in Table 5. The calibration validity between CLQ, MSLQ, and SCLS was analyzed based on a valid sample size of 376. In CLQ, the SCLS total score was correlated with 16 mental effort items, of which 15 items were significantly correlated ($0.22 \leq r \leq 0.66$, $p < 0.01$); there was no significant correlation between autonomous input of SCLS and emotion engagement of mental effort. In the MLSQ (Chinese version), SCLS has 12 correlations with autonomous input, all of which are significantly correlated ($0.27 \leq r \leq 0.83$, $p < 0.001$). In academic scores, the SCLS total score was significantly positively correlated with the academic total score and English score ($0.10 \leq r \leq 0.13$, $p < 0.05$) and had no significant correlation with math or Chinese scores. There was no significant correlation between the environmental coping and the autonomous input with the scores of all subjects.

TABLE 5 Partial correlations between SCLS (and its three subscales), CLQ, MLSQ, and academic scores.

Variables	Sample (*n*)	Awareness	Input	Coping	Total CL
CLQ	2 (*150*)	0.60***	0.27***	0.35***	0.54***
Mental	2 (*150*)	0.66***	0.22**	0.38***	0.55***
Emotion	2 (*150*)	0.35***	0.12	0.22**	0.30***
Time	2 (*150*)	0.43***	0.30***	0.23**	0.42***
MLSQ	3 (*226*)	0.83***	0.45***	0.32***	0.76***
Strategy Use	3 (*226*)	0.76***	0.36***	0.27***	0.66***
Self-Regulation	3 (*226*)	0.76***	0.49***	0.32***	0.74***
Academic Scores	2 + 3 (*376*)	0.19***	0.02	(0.01)	0.10*
Chinese	2 + 3 (*376*)	0.18***	(0.00)	0.01	0.10
Math	2 + 3 (*376*)	0.11*	0.04	(0.04)	0.06
English	2 + 3 (*376*)	0.24***	0.02	(0.01)	0.13*

Awareness = learning awareness; input = autonomous input; coping = environmental coping; MLSQ = Motivated Strategies for Learning Questionnaire; CLQ = Questionnaire of Cognitive Load; mental = mental engagement; emotion = emotion engagement; time = time engagement. Partial correlations in sample 2 ($n = 150$) and sample 3 ($n = 226$) after removing variances of gender, grade, and age; partial correlations in sample 2 with sample 3 ($n = 376$) after removing variances of school level, gender, grade, and age.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Discussion

The purpose of this research was to create and validate a survey measuring student conceptual level, specifically for secondary school students. The EFA of SCLS showed that the statistical results were consistent with the constructed theoretical framework of conceptual level and that the three factors both independently represented the significance of their respective scores and combined to reflect the overall conceptual level. The three factors contributed 51.46% of the total variance. The results of the CFA test showed that the model had high structural validity and met the assumptions of the empirical construction of students' conceptual level. Reliability and validity analyses showed high correlations between the dimensions of the scale and the variables. The Cronbach's α, Spearman–Brown, CR, and retest reliability of both the SCLS total and subscales exceeded the 0.6 critical standard, indicating good overall internal consistency and retest reliability. The results of this study provide sufficient validity evidence for the scores, generalizations, and extrapolations of the SCLS' full scale and subscales, as well as for the significance and utility of the subscales, indicating that the scale can assess student CL.

Empirical construction of conceptual level

In the social domain, CL is defined as a personality trait. In the teaching field, CL is defined as learning style. Both define CL as a developmental index that has value for personal growth (Harvey et al., 1962; Hunt et al., 1978). In addition to affirming the value of CL, there is some disagreement and confusion between the two fields as to the exact meaning and internal structure of CL. First, the CL for personality orientation describes only cognitive dispositions, whereas the CL for learning style orientation describes both cognitive dispositions and behavioral characteristics. Second, the CL for personality orientation is how individuals behave in general, whereas the CL for learning style orientation is how students behave in learning situations. Finally, the initial measurement scale for this study adapted the PCM themes about rules, authority, and parenting but did not show good statistical properties and was solidly removed. The above three points show that the theoretical connotation and internal structure of the CL of learning style orientation have changed; not only the connotation is different from the personality orientation, but also the measurement instrument proves the inapplicability of PCM. This study explores the changes in the theoretical connotation and inner structure of CL with learning style orientation, not only theoretically deepening the learning style connotation of CL but also addressing the reality of measurement tools.

In quantitative research, a clear definition of a concept is the basis for operationalizing and quantifying that concept. Based on the connotations of learning in CL, this study is empirically constructed using a realist perspective around the characteristics of the three aspects proposed by Hunt (cognitive complexity, independence, and environmental adaptability) and analyzes which characteristics students have that reflect these three aspects and which specific factors these three aspects focus on. Based on the theoretical constructs, this study used logical methods such as generalization and induction to gradually construct the empirical dimensions of students' CL.

Item screening was a key component in the development of the SCLS. To improve the reliability of the scale, this study analyzed each item in the scale through item analysis to find differences between high and low subgroups and to screen out high-quality items to constitute the scale. The correlation coefficients between the scale items and the full scale reached statistically significant levels, except for the expert evaluations. The correlation between each item and the total score was significant, with coefficients ranging from 0.30 to 0.64. The skewness of all items ranged from −0.953 to 0.426, which did not exceed the critical value of 3. The kurtosis ranged from −1.287 to 0.829, which was not >10, indicating that the scores of all items conformed to a normal distribution and had a high degree of discrimination.

Dimensions of conceptual level

The student conceptual level consists of three factors—learning awareness, autonomous input, and environmental coping—reflecting the three aspects proposed by Hunt: cognitive complexity, independence, and environmental adaptability. These three relatively independent but interrelated components make up the student conceptual level.

Among them, learning awareness reflects the level of development of students on the concrete–abstract cognitive dimension of learning itself, of their own learning state. For example, for students with low learning awareness, teachers should initially set goals and plans for them, organize simpler learning tasks, and pay attention to how students' cognition evolves on the concrete–abstract dimension. When students' cognitive abstraction level is improved, teachers should reduce participation and encourage students to plan their own learning. Autonomous input reflects the level of student development on the dependent–independent learning dimension. For example, for students with high levels of dependency, teachers or parents should act as "supervisors" and try to create a learning environment free of distractions and temptations. When students become more independent, teachers or parents should provide a free learning environment, reduce intervention, and encourage students to self-manage. Environmental coping reflects the level of students' processing of external events on the conflict–balance dimension. For example, when students with a low level of environmental coping encounter conflict or difficult situations (interpersonal and learning situations), teachers should promptly help them construct a good psychological environment and develop interpersonal harmony and psychological resilience.

The above three factors validate the initial theoretical constructs of this study, and the meaning of each is well defined. Therefore, the questionnaire structure of this study is complete.

Reliability and validity of the SCLS

In this study, construct validity, discriminant validity, and criterion validity were used to reflect the overall validity of the scale.

CFA showed that the three-factor structural model of SCLS obtained by EFA had a good fit index. Significant correlations ($p < 0.001$) were found among the factors and between each factor and the total score, indicating that the scale has good construct validity.

The AVE' square root for all three factors was greater than the correlation coefficients among the three, indicating good discriminant validity.

In this study, two internal criteria were selected to reflect concurrent validity. Among them, LCQ reflected mental effort. Sixteen of the SCLS items were correlated with the LCQ, 15 of which were significant, confirming a strong relationship between mental effort and individual characteristics (Miller, 1956; Sweller, 2010; Sweller et al., 2019; Paas and Merrinboer, 2020; Brosnan et al., 2021). However, the correlation between autonomous input and emotional engagement (controlling bad emotions and regulating bad moods) was not significant, probably because the level of autonomous input involved cognitive and behavioral dimensions, whereas emotional engagement reflected emotional management, and emotions were not always consistent with cognition and behavior. The MSLQ (Chinese) reflects self-directed learning ability. There were 12 item correlations between SCLS and MSLQ (Chinese), all of which were significant ($0.27 \leq r \leq 0.83$, $p < 0.001$). Research has found that students who use effective learning strategies and have high levels of self-management are more likely to learn autonomously and have a more internal and deeper understanding of the meaning of learning (Zimmerman, 1986; Pintrich, 2000; Efklides et al., 2018). This supports the theoretical connotation of CL as a developmental goal of "autonomy": It reflects the degree of autonomy of the learner (Harvey et al., 1962; Hunt, 1971, 1975a, 1976a). The above two validity tests demonstrated the good concurrent validity of the SCLS.

This study also selected students' academic scores as an external validity marker to reflect predictive validity. Total SCLS scores were significantly positively correlated with total academic scores and English scores, while all other factors were not significantly correlated. This result aptly demonstrates the theoretical connotation of CL: CL is a learning style (Hunt, 1971; Hunt et al., 1974; Tan, 1995; Flowers et al., 2000) that reflects how one learns, not what one learns. Academic scores are a test of accumulated knowledge experience, reflecting what was learned, not how it was learned. CL cannot be linked to academic scores without matching with instruction, and only matched instruction can truly facilitate student learning, which is the core meaning of CL in teaching and learning (Hunt et al., 1974, 1978; Miller, 1981; Tan, 1995). Thus, CL is not simply linearly related to academic scores, and the relationship should be interpreted with caution (Miller, 1978). Specifically, although the total SCLS score was significantly positively correlated with both total academic scores and English scores ($0.11 \leq r \leq 0.24$, $p < 0.05$), the correlation was low. The total SCLS score was not significantly correlated with math and Chinese scores. Only the learning awareness was significantly positively correlated with scores in all subjects ($0.11 \leq r \leq 0.24$, $p < 0.05$). It is not difficult to explain that learning awareness reflects students' in-depth thinking and execution of learning, while strategies and planning predict the achievement of course goals (Kizilcec et al., 2017). Autonomous input was not significantly correlated with all subjects' scores. This can be explained in two ways: (1) The relationship between input and academic scores is not unidirectional; high input does not mean a high score (Bruin et al., 2020). For example, students can complete academic work and perform well and get good scores without necessarily being invested in it; conversely, there are students who invest a lot of effort in some details and procedures but do not really develop their thinking. This has been classified into two types of effort: objective effort and subjective effort (Dunlosky and Mueller, 2016; Dunlosky et al., 2020). The former is a superficial input, while the latter is a substantive input, and only a substantive input can truly facilitate learning. (2) From the perspective of economics, there is a law of diminishing marginal returns in the learning process: With the increase of learning input, the initial achievement increases, and when the input exceeds a certain limit, this increasing trend gradually decreases and eventually decreases absolutely (Bai, 1999; Shen, 2014), which indicates that the increase of learning input does not necessarily lead to the improvement of learning achievement, and the key is to find the optimal combination between learning inputs and outputs. Environmental coping was not significantly correlated with all subjects' scores. It reflects how students balance themselves with various elements of their environment and is an unrelated aspect of learning; therefore, it does not directly reflect achievement and may be influenced by other factors. The relationship between each subject score and the total SCLS score needs to be analyzed according to the subject characteristics of the secondary school (grade7-9) curriculum of the Ministry of Education (2011). Chinese and math scores reflect thinking levels and prior knowledge more prominently than English scores and require higher levels of learning ability. Meanwhile, CL is not a learning ability but a cognitive and behavioral characteristic. Secondary school (grade7-9) English does not require a high level of thinking and *a priori* knowledge and it is still at the primary stage. English learning highlights individual differences in learning characteristics, habits, language communication, social interaction, etc. This is consistent with the connotation of individual differences reflected by CL, so the association between the subject of English and CL is relatively close.

All reliability indicators, such as Cronbach's α, retest reliability, and Spearman–Brown, met the acceptable criteria for reliability coefficients in this study (Wu, 2018, P244). However, Cronbach's α for the subscales was only at or near the critical value. This may be related to the length of the test: The number of items on each factor of the SCLS was 3 or 4, which may have led to low reliability. An effective solution would be to increase the length of the test appropriately (Dai et al., 1999).

 10.3389/feduc.2022.965643

Limitations and future directions

Owing to the slow research progress of CL theory in the past 20 years, this study focused on the classic literature from the 1960s–1980s in terms of literature acquisition and reference. When we were reconstructing this theory, we lacked empirical comparative studies in the context of the changing times. At the same time, there may be some differences between the grasping of the meaning of the classical theory and the content freely generated by the contemporary subjects. Of course, the theory in these classic texts is not cut off from later generations in the study of personality traits, and its successful application in the field of counseling and therapy is sufficient to show that CL has had a profound influence on personality research. The present study does not further follow the classical theoretical ideas to find a broader developmentally appropriate ground for CL and only explores the significance and internal structure of CL in terms of learning styles, without restoring its significance in personality development and re-exploring the internal structure. Therefore, it is difficult to answer the question, "Has CL also undergone fundamental changes in the assessment of personality based on the social domain?" in this study. In general, the secondary school level is only at a certain stage of learning, with particular temporal and environmental characteristics, while individual learning is permeated by various environments and stages of existence in question. Therefore, it is necessary to explore aspects of CL that may be of value in multiple contexts and stages.

Although the results of this study help to define the connotation of learning styles in CL, there are still some issues to be addressed in terms of validity. Additional follow-up research is needed to address the limitations of this study. (1) Although the characteristics of the sample can be generalized to populations with similar characteristics, they may not apply to populations with different characteristics, lifestyles, or ages in other countries. In future studies, it will be necessary to consider samples from other cultures and test for invariance across groups to extend and support the findings and contributions of this study. (2) It is necessary to combine other forms, such as individual qualitative analysis techniques (e.g., semi-structured interviews, observation, content analysis of autobiographical texts) to deepen the nature of CL authenticity and individual differences among students.

Conclusion

The SCLS developed in this study has good reliability and validity, meets psychometric requirements, and reflects the actual status of the students' CL. The SCLS has 13 items, including 3 factors: learning awareness, autonomous input, and environmental coping. This study demonstrated that the Student Conceptual Level Scale (SCLS) provides an efficient, reliable, and accurate way to assess learning styles. The SCLS can be used not only to assess student learning and individual development but also to help us understand the causes of differences in CL across students and to reduce this gap through instruction.

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.

Ethics statement

The studies involving human participants were reviewed and approved by Nanjing Normal University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

XY was responsible for the original conception, data collection, and data analysis and constructed the original concepts. DT and JD critically revised the manuscript. All authors approved the final version of the manuscript for submission.

Acknowledgments

We are grateful for the generous contributions of study participants and staff who assisted with data collection during the study. We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this 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.

References

Bai, Y. M. (1999). A review of the relationship between learning time and learning outcome. *Stud. Foreign Educ.* 6:7.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246. doi: 10.1037/0033-2909.107.2.238

Brill, R. (1978). Implications of the cl matching model for treatment of delinquents. *J. Res. Crime Delinq.* 15, 229–246. doi: 10.1177/002242787801500208

Brophy, J. E., and Good, T. L. (1974). Teacher-student relationships: causes and consequences. *Am. J. Educ.* 6, 249–250. Available at: https://eric.ed.gov/?id=ED091495

Brosnan, K., Grün, B., and Dolnicar, S. (2021). Cognitive load reduction strategies in questionnaire design. *Int. J. Mark. Res.* 63, 125–133. doi: 10.1177/1470785320986797

Browne, M. W., and Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociol. Methods Res.* 21, 230–258. doi: 10.1177/0049124192021002005

Bruin, A. B. H. D., Roelle, J., Carpenter, S. K., Baars, M., and Trypke, M. (2020). Synthesizing cognitive load and self-regulation theory: a theoretical framework and research agenda. *Educ. Psychol. Rev.* 32, 903–915. doi: 10.1007/s10648-020-09576-4

Carpenter, S. K., Endres, T., and Hui, L. (2020). Students' use of retrieval in self-regulated learning: implications for monitoring and regulating effortful learning experiences. *Educ. Psychol. Rev.* 32, 1029–1054. doi: 10.1007/s10648-020-09562-w

Chang, C. C., Liang, C. Y., Chou, P. N., and Lin, G. Y. (2017). Is game-based learning better in flow experience and various types of cognitive load than non-game-based learning? Perspective from multimedia and media richness. *Comput. Hum. Behav.* 71, 218–227. doi: 10.1016/j.chb.2017.01.031

Curran, P. G. (2015). Methods for the detection of carelessly invalid responses in survey data. *J. Exp. Soc. Psychol.* 66, 4–19. doi: 10.1016/j.jesp.2015.07.006

Dai, H. Q., Zhang, F., and Chen, X. F. (1999). *Psychological and Educational Measurement.* (Jinan: Jinan University Press). 241–242.

Doyle, W., and Rutherford, B. (1984). Classroom research on matching learning and teaching styles. *Theory Pract.* 23, 20–25. doi: 10.2307/1476734

Dunlosky, J., Badali, S., Rivers, M. L., and Rawson, K. A. (2020). The role of effort in understanding educational achievement: objective effort as an explanatory construct versus effort as a student perception. *Educ. Psychol. Rev.* 32, 1163–1175. doi: 10.1007/s10648-020-09577-3

Dunlosky, J., and Mueller, M. (2016). Recommendations for exploring the disfluency hypothesis toward establishing whether perceptually degrading materials impacts performance. *Metacogn. Learn.* 11, 123–131. doi: 10.1007/s11409-016-9155-9

Dunn, A. M., Heggestad, E. D., Shanock, L. R., and Theilgard, N. (2016). Intra-individual response variability as an indicator of insufficient effort responding: comparison to other indicators and relationships with individual differences. *J. Bus. Psychol.* 33, 105–121. doi: 10.1007/s10869-016-9479-0

Efklides, A., Schwartz, B. L., and Brown, V. (2018). "Motivation and affect in self-regulated learning: does metacognition play a role," in *Handbook of Self-regulation of Learning and Performance. 2nd Edn.* eds. D. H. Schunk and J. A. Greene.

Erikson, E. H. (1959). Identity and the life cycle: selected papers. *Psychological Issues*. (US: International Universities Press, Inc.), 1–171.

Fan, J. Y., Ye, B., Zhang, Z. Y., and Liu, B. X. (2003). Exploratory factor analysis: a literature review between 1991 and 2000. *Adv. Psychol. Sci.* 11, 579–585. doi: 10.3969/j.issn.1671-3710.2003.05.018

Fang, J. M., Prybutok, V., and Wen, C. (2016). Shirking behavior and socially desirable responding in online surveys: a cross-cultural study comparing Chinese and American samples. *Comput. Hum. Behav.* 54, 310–317. doi: 10.1016/j.chb.2015.08.019

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics (Paperback and Web Assign). 4th Edn*. SAGE Publications Ltd.

Flowers, C. P., Hancock, D. R., and Joyner, R. E. (2000). Effects of instructional strategies and cls on students' motivation and achievement in a technology course. *J. Res. Dev. Educ.* 33, 187–194. Available at: https://eric.ed.gov/?id=EJ612234

Gardiner, G. S., and Schroder, H. M. (1972). Reliability and validity of the paragraph completion test: theoretical and empirical notes. *Psychol. Rep.* 31, 959–962. doi: 10.2466/pr0.1972.31.3.959

Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., and Eysink, T. (2009). Learning with hypermedia: the influence of representational formats and different levels of learner control on performance and learning behavior. *Comput. Hum. Behav.* 25, 360–370. doi: 10.1016/j.chb.2008.12.015

Ghazivakili, Z., Nia, R. N., Panahi, F., Karimi, M., and Ahmadi, Z. (2014). The role of critical thinking skills and learning styles of university students in their academic performance. *J. Adv. Med. Educ. Prof.* 2, 95–102. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4235550/pdf/jamp-2-95.pdf.

Goldberg, A. D. (1974). Conceptual system as a predisposition toward therapeutic communication. *J. Couns. Psychol.* 21, 364–368. doi: 10.1037/h0037108

Gravetter, F. J., and Wallnau, L. B. (2014). *Essentials of Statistics for the Behavioral Sciences 8th Edn.* Available at: https://xueshu.baidu.com/usercenter/paper/show?paperid=ed7fae97af6284c7ad70df26587704ac

Greaves, G. (1971). Harvey's "this I believe" test: studies of reliability. *Psychol. Rep.* 28, 387–390. doi: 10.2466/pr0.1971.28.2.387

Harold, M. S., Michael, J. D., and Siegfried, S. (1967). Human information processing: individuals and groups functioning in complex social situations. *Am. J. Psychol.* 83:136. doi: 10.2307/1420865

Harvey, O. J., Hunt, D. E., and Schroder, H. M. (1962). Conceptual systems and personality organization. *Psychosomatics* 76:163. doi: 10.2307/1420024

Hayton, J. C., Allen, D. G., and Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organ. Res. Methods* 7, 191–205. doi: 10.1177/1094428104263675

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., and DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology* 27, 99–114. doi: 10.2307/41474909

Hunt, D. E. (1970). A CL matching model for coordinating learner characteristics with educational approaches. *Interchange* 1, 68–82. doi: 10.1007/BF02214677

Hunt, D. E. (1971). *Matching Models in Education: The Coordination of Teaching Methods with Student Characteristics*, Toronto: Ontario Institute for Studies in Education. Available at: https://psycnet.apa.org/record/1971-27828-001

Hunt, D. E. (1975a). Person-environment interaction: a challenge found wanting before it was tried. *Rev. Educ. Res.* 45, 209–230. doi: 10.3102/00346543045002209

Hunt, D. E. (1975b). The B-P-E paradigm for theory, research and practice. *Can. Psychol. Rev. Psychologie Canadienne* 16, 185–197. doi: 10.1037/h0081805

Hunt, D. E. (1975c). Teachers' adaptation to students: implicit and explicit matching. Research and development memorandum no. 139. *Adjustment*:50. Available at: https://eric.ed.gov/?id=ED117090

Hunt, D. E. (1976a). Teachers are psychologists, too: on the application of psychology to education. *ACT Res. Rep.* 17, 210–218. doi: 10.1037/h0081840

Hunt, D. E. (1976b). Teachers' adaptation: 'Reading' and 'Flexing' to students. *J. Teach. Educ.* 27, 268–275. doi: 10.1177/002248717602700323

Hunt, D. E. (1977). Conceptual level theory and research as guides to educational practice. *Interchange* 8, 78–90. doi: 10.1007/BF01810464

Hunt, D. E. (1979). "Learning style and student needs: an introduction to conceptual level," in *Student Learning Styles: Diagnosing and Prescribing Programs*. ed. NASSP (Reston: National Association of High School Principals), 27–38. Available at: https://eric.ed.gov/?id=ED182859

Hunt, D. E., Butler, L. F., Noy, J. E., and Rosser, M. E. (1978). *Assessing CL by the Paragraph Completion Method*. Toronto: Ontario Institute for Studies in Education. (unpublished online).

Hunt, D. E., Joyce, B. R., Greenwood, J. A., Noy, J. E., Reid, R., and Weil, M. (1974). Student cl and models of teaching: theoretical and empirical coordination of two models. *Interchange* 5, 19–30. doi: 10.1007/BF02138804

Hutcheson, G. D., and Sofroniou, N. (1999). *The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models.* Sage Publications.

İlçin, N., Tomruk, M., Yeşilyaprak, S. S., Karadibak, D., and Savcı, S. (2018). The relationship between learning styles and academic performance in Turkish physiotherapy students. *BMC Med. Educ.* 18:291. doi: 10.1186/s12909-018-1400-2

Kizilcec, R. F., Pérez-Sanagustín, M., and Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Comput. Educ.* 104, 18–33. doi: 10.1016/j.compedu.2016.10.001

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., and Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* 45, 1058–1072. doi: 10.3758/s13428-013-0334-1

Lewin, K. (1936). A dynamic theory of personality: selected papers. *J. Nerv. Ment. Dis.* 33, 548–552. doi: 10.1037/h0050556

Li, X. D., and Lin, C. D. (2001). Academic help-seeking strategies: its relation to personal goal orientation. *Classroom Goal Struct. Culture Factors Psychol. Dev. Educ.* 17:7. doi: 10.3969/j.issn.1001-4918.2001.02.001

Li, H. M., and Luo, S. Q. (2014). Teaching efficacy logic of academic burden problem solving. *Res. Educ. Dev.* 10:6. Available at: http://en.cnki.com.cn/Article_en/CJFDTOTAL-SHGJ201410016.htm

Li, Z. J., and Yin, H. B. (2010). The impact of classroom environment on Students' self-regulated learning in Hong Kong: a commentary on the debate of teacher-centeredness and student-centeredness. *Peking Univ. Educ. Rev.* 1, 70–82. doi: 10.19355/j.cnki.1671-9468.2010.01.008

Lin, T. J., Deng, F., Chai, C. S., and Tsai, C. C. (2013). High school students' scientific epistemological beliefs, motivation in learning science, and their relationships: a comparative study within the Chinese culture. *Int. J. Educ. Dev.* 33, 37–47. doi: 10.1016/j.ijedudev.2012.01.007

McLachlan, J. F. (1974). Therapy strategies, personality orientation and recovery from alcoholism. *Can. Psychiatr. Assoc. J.* 19, 25–30. doi: 10.1177/070674377401900106

McLachlan, J. F., and Hunt, D. E. (1973). Differential effects of discovery learning as a function of student CL. *Can. J. Behaviouralence Revue Canadienne Desences Du Comportement* 5, 152–160. doi: 10.1037/h0082340

Miller, G. A. (1956). The magic number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 101, 343–352. doi: 10.1037/0033-295X.101.2.343

Miller, A. (1978). Conceptual systems theory: a critical review. *Genet. Psychol. Monogr.* 97, 77–126.

Miller, A. (1981). Conceptual matching models and interactional research in education. *Rev. Educ. Res.* 51, 33–84. doi: 10.3102/00346543051001033

Ministry of Education (2011). *Curriculum Standards for Compulsory Education*. Beijing: PRC. Available at: http://www.moe.gov.cn/srcsite/A26/s8001/201112/t20111228_167340.html (Accessed May 30, 2022).

Murphy, P. D., and Brown, M. M. (1970). Conceptual systems and teaching styles. *Am. Educ. Res. J.* 7, 529–540. doi: 10.2307/1161834

Nie, Y. G., and Ding, L. (2009). The characters of adolescent self-consciousness and its relationship with social adaptive behavior. *Psychol. Dev. Educ.* 2:8. doi: 10.3969/j.issn.1674-196X.2014.37.043

O'connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and velicer's map test. *Behav. Res. Methods Instrum. Comput.* 32, 396–402. doi: 10.3758/BF03200807

Paas, F., and Merrinboer, J. J. G. V. (2020). Cognitive-load theory: methods to manage working memory load in the learning of complex tasks. *Curr. Dir. Psychol. Sci.* 29, 394–398. doi: 10.1177/0963721420922183

Pavlov, G., Maydeu-Olivares, A., and Shi, D. (2021). Using the standardized root mean squared residual (SRMR) to assess exact fit in structural equation models. *Educ. Psychol. Meas.* 81, 110–130. doi: 10.1177/0013164420926231

Phillips, M., and Sinclair, R. (1973). Conceptual systems and educational environment: relationships between teacher conceptual systems, student conceptual systems, and classroom environment as perceived by fifth and sixth grade students. *Classroom Environ.* 30, 1–23. Available at: https://files.eric.ed.gov/fulltext/ED076539.pdf

Piaget, J. (1953). The child's conception of number. *American Sociological Review*. 18, 711–712. doi: 10.2307/2088144

Pintrich, P. R. (2000). "The role of goal orientation in self-regulated learning," in *Handbook of Self-regulation*. eds. M. Boekaerts, P. R. Pintrich and M. Zeidner, 451–494. doi: 10.1016/B978-012109890-2/50043-3

Rao, N., and Sachs, J. (1999). Confirmatory factor analysis of the Chinese version of the motivated strategies for learning questionnaire. *Educ. Psychol. Meas.* 59, 1016–1029. doi: 10.1177/00131649921970206

Shen, X. J. (2014). Research on Shanghai high school Students' optimal learning time based on PISA data. *Res. Educ. Dev.* 4:6. doi: 10.14121/j.cnki.1008-3855.2014.04.014

Stoppard, J. M., and Miller, A. (1985). Conceptual level matching in therapy: a review. *Current Psychol. Res. Rev.* 4, 46–68. doi: 10.1007/BF02686567

Sun, C. Y. (2016). Research on adolescent academic burden from the perspective of cognitive load. *Educ. Exploration*, 4. Available at: http://www.cnki.com.cn/Article/CJFDTotal-SEEK201604010.htm

Sun, C. Y., and Liu, D. Z. (2016). Effects of the background color of learning materials on cognitive load and learning performance. *J. Psychol. Sci.* 39, 869–874. doi: 10.16719/j.cnki.1671-6981.20160416

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* 22, 123–138. doi: 10.1007/s10648-010-9128-5

Sweller, J., van Merrienboer, J. J. G., and Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* 31, 261–292. doi: 10.1007/s10648-019-09465-5

Tan, D. L. (1995). *Learning style treatise* (Nanjing: Jiangsu Education Press). 9–, 111–123.

Tomlinson, P. D., and Hunt, D. E. (1971). Differential effects of rule-example order as a function of learner conceptual level. *Can. J. Behav. Sci./revue Canadienne Des Sciences Du Comportement* 3, 237–245. doi: 10.1037/h0082265

Trochim, W. M., and Donnelly, J. P. (2008). *The Research Methods Knowledge Base*. Available at: https://www.researchgate.net/publication/281345819_The_Research_Methods_Knowledge_Base (Accessed May 30, 2022).

Tucker, L. R., and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* 38, 1–10. doi: 10.1007/BF02291170

Tuckman, W. B. (1965). Integrative complexity: its measurement and relation to creativity. *Educ. Psychol. Meas.* 26, 369–382. doi: 10.1177/001316446602600209

Wang, H. J., and Lu, J. M. (2004). The compilation of the high school students' self-control ability questionnaire. *J. Psychol. Sci.* 6, 198–203. doi: 10.16719/j.cnki.1671-6981.2004.06.055

Wirth, J., St Ebner, F., Trypke, M., Schuster, C., and Leutner, D. (2020). An interactive layers model of self-regulated learning and cognitive load. *Educ. Psychol. Rev.* 32, 1127–1149. doi: 10.1007/s10648-020-09568-4

Wong, J., Baars, M., Davis, D., Tim, V. D. Z., Houben, G. J., and Paas, F. (2018). Supporting self-regulated learning in online learning environments and moocs: a systematic review. *Int. J. Hum.-Comput. Interact.* 35, 356–373. doi: 10.1080/10447318.2018.1543084

Wu, M. L. (2018). *Practice of questionnaire statistical analysis: Operation and application of SPSS.* (Chongqing: Chongqing University Press). 244.

Yazici, H. J. (2016). Role of learning style preferences and interactive response systems on student learning outcomes. *Int. J. Inf. Oper. Manage. Educ.* 6:109. doi: 10.1504/IJIOME.2016.076046

Zhang, Q., Wang, Y., Zhang, L., and Zhang, H. (2010). The development of high school students' informal reasoning and influence factors. *Psychol. Explor.* 6:6. doi: 10.3969/j.issn.1003-5184.2010.06.013

Zhao, J.-F. (2011). *Deciphering academic load: A study of cognitive load in learning.* (Beijing: Science Press). 85–86.

Zhou, H., and Long, L. R. (2004). Statistical remedies for common method biases. *Adv. Psychol. Sci.* 12:9. doi: 10.3969/j.issn.1671-3710.2004.06.018

Zimmerman, B. J. (1986). Becoming a self-regulated learner: which are the key subprocesses? *Contemp. Educ. Psychol.* 11, 307–313. doi: 10.1016/0361-476X(86)90027-5

frontiers | Frontiers in Education

TYPE Methods
PUBLISHED 09 February 2023
DOI 10.3389/feduc.2023.1087908

Check for updates

OPEN ACCESS

EDITED BY
Antonio P. Gutierrez de Blume,
Georgia Southern University,
United States

REVIEWED BY
Vahid Nimehchisalem,
Putra Malaysia University,
Malaysia
Kathryn Holmes,
Western Sydney University,
Australia

*CORRESPONDENCE
Matthias Borgstede
✉ matthias.borgstede@uni-bamberg.de

SPECIALTY SECTION
This article was submitted to
Teacher Education,
a section of the journal
Frontiers in Education

RECEIVED 02 November 2022
ACCEPTED 18 January 2023
PUBLISHED 09 February 2023

CITATION
Borgstede M and Rau C (2023) Beyond quality and quantity: Representing empirical structures by embedded typologies.
Front. Educ. 8:1087908.
doi: 10.3389/feduc.2023.1087908

Beyond quality and quantity: Representing empirical structures by embedded typologies

Matthias Borgstede* and Caroline Rau

Foundations of Education, University of Bamberg, Bamberg, Germany

In this article, we propose a new approach to the problem of integration in mixed methods research that builds on a representational understanding of empirical science. From this perspective, qualitative and quantitative modeling strategies constitute two different ways to represent empirical structures. Whereas qualitative representations focus on the construction of *types* from cases, quantitative representations focus on the construction of *dimensions* from variables. We argue that types and dimensions should be integrated within a joint representation of the data that equally acknowledges qualitative and quantitative aspects. We outline how the proposed representational framework can be used to embed qualitative types in quantitative dimensions using an empirical study on teachers' epistemological beliefs.

KEYWORDS

qualitative and quantitative research, mixed methods, documentary method, discriminant analysis, cluster analysis, teachers' epistemological beliefs

1. Introduction

Although qualitative and quantitative research methods adhere to rather distinct assumptions about the aims and scope of scientific enquiry (Freeman et al., 2007), attempts to integrate both strategies within a mixed methods approach are eventually gaining attention (Creswell, 2015). Mixed methods research attempts to combine the strengths of qualitative and quantitative research by using both strategies within an overarching methodological framework. However, there is no generally accepted integration framework for mixed methods designs (Fetters and Molina-Azorin, 2017). Consequently, integration of qualitative and quantitative strategies remains a major challenge for mixed methods research (Moran-Ellis et al., 2006; Creswell, 2009; O'Cathain et al., 2010; Fielding, 2012; Bazeley, 2016). In some cases, the qualitative and quantitative parts of a mixed methods design may even appear to be two different studies that are only connected thematically (Yin, 2006; Bergman, 2011).

In this article, we approach the problem of integration from a representational perspective that characterizes qualitative and quantitative research as different ways to represent empirical observations by means of scientific theories (Borgstede and Scholz, 2021). In this view, the semantics of a scientific theory (or model) is twofold. On the one hand, it consists of the *meaning* of its constituting concepts. On the other hand, it consists of the *topology* that relates these concepts to one another. Different types of models may emphasize either one of these aspects. Following Borgstede and Scholz (2021), qualitative representations focus on theoretically meaningful concept formation that is based on the essential properties of the objects under study. The objects may be of very different nature, depending on the focus of the study. This includes inanimate objects, as well as living organisms, people, groups, interviews, or even abstract concepts and themes. Quantitative representations, however, are more concerned with the topological structure that relates the concepts to one another. In other words, they focus on the dimensions by which the concepts differ and their geometric properties.

We argue that acknowledging that meaning and topology are just different aspects of the semantics of scientific concepts is the key to a general framework for integrating qualitative and quantitative research. Qualitative research captures the essential properties of scientific concepts by means of abstract typologies. Quantitative research captures the relations between different scientific concepts by means of geometric spaces.

We develop a corresponding methodological framework that integrates both perspectives to answer the following questions: (a) How do qualitative representations (typologies) relate to quantitative representations (geometric spaces)? (b) How can the former be properly embedded into the latter? The focus of our analysis is a common research problem that arises when the results of a qualitative study are used to guide quantitative model building and test construction. Nevertheless, the methodology can easily be adapted to other mixed methods designs. For example, a study may start with quantitative questionnaire data and then proceed with a qualitative investigation to further explore the underlying empirical structure. Once the qualitative data has been collected, our framework may be applied in exactly the same way as if the qualitative study had not been informed by questionnaire data. In addition to the quantitative embedding of qualitative types, the resulting geometric space may be used to revise the original questionnaires and inform further quantitative inquiry.

In the following sections, we will first elaborate on the general relation between (qualitative) case-based and (quantitative) variable-based models from a representational perspective (section 2). We will proceed to characterize both approaches with an emphasis on explicit and implicit model properties (sections 3, 4). We will then refine the representational view to develop a methodological framework for integrating qualitative and quantitative research (section 5), and demonstrate the feasibility of our framework by an empirical case study about teachers' epistemological beliefs (section 6). Finally, we discuss the implications of our analysis with regard to further methodological developments and possible applications (section 7).

2. Qualitative and quantitative representations

Borgstede and Scholz (2021) argue that qualitative and quantitative research employ two different, yet compatible, ways to describe empirical relational structures. Whereas the qualitative strategy uses a *case-based* approach to characterize individuals, the quantitative strategy applies a *variable-based* approach to characterize attributes and their functional relations (Ragin, 1987; Rihoux and Ragin, 2009). In many cases, however, both modeling strategies can be applied to one and the same empirical structure. Consequently, qualitative (case-based) models often imply a quantitative structure that is distinct from the observed qualia. Similarly, quantitative (variable-based) models often imply a qualitative structure that is distinct from the variables that are used to represent the data.

To illustrate the relation between qualitative and quantitative representations, let us imagine a simple empirical structure consisting of two kinds of objects—a collection of cases and a collection of observations. The cases may be individuals, groups, texts or instances of any other category. The observations may be distinct behaviors, answers in a standardized test, utterances in a conversation or any other class of attributes of the cases under study.[1]

If the cases are individuals and the observations are the answers in a competence test, we might represent the observed empirical structure by means of a quantitative model involving a single dimension. One of the most commonly used models for such structures is the Rasch model (Rasch, 1960). The Rasch model postulates a single quantitative dimension by which individuals and test items can be compared, such that higher differences between an individual and a test item result in a higher probability of a correct answer. This probabilistic relation makes some empirical structures (i.e., answer patterns in a test) more likely than others. Hence, if an observed empirical structure is not too unlikely given the Rasch model, we can use the structure to statistically estimate the corresponding model parameters. This scaling procedure transforms the empirical structure that consisted of individuals and test items into a new abstract structure that consists of points in a unidimensional geometric space – a so-called latent variable. We have thus constructed a numerical representation of an empirical structure. Such representations are called *measurement* (Krantz et al., 1971). Measurement, as described above, is the foundation of any quantitative science. However, it is important to note that proper measurement has to be grounded in suitable empirical structures.[2]

If the cases are, for example, work teams and the observations are utterances in a discussion, we would rarely apply a Rasch model to represent the empirical structure (although, technically, this would be possible). In such a case the more intuitive approach would be to group the different utterances according to their semantic similarity and to group the teams such that they maximally differ with respect to the semantic content of their utterances. Instead of mathematical model equations and statistical estimation techniques, such a grouping usually relies on interpretative acts from the part of the researchers. However, the result is also a representation of the empirical structure, the main difference being that instead of numbers we have used words, or abstract concepts, to guide our interpretation of the topics and themes that have been discussed by the work teams under study. The interpretative act transforms the empirical structure that consisted of teams and utterances into a new abstract structure that consists of categories and types. We have constructed a conceptual representation of an empirical structure. Such representations are commonly called *typologies* or, if the grouping is only performed over the topics rather than the cases, *category systems* or *patterns*. Typologies and category systems are the foundation of all qualitative science. Like measurement, such representations have to be grounded in empirical observations (Flick, 2014). In light of these two examples, quantitative and qualitative approaches seem not so different after all. Both rely on empirical structures consisting of cases and observations. And both provide means to construct abstract representations of these empirical structures. However, categories are not dimensions and typologies are not geometric spaces.

Qualitative and quantitative research can be consistently interpreted as specific modeling strategies. They may even both be applied to one

1 Mathematically, cases and observations are two different kinds of *objects* that jointly form a *relational structure*.

2 In fact, there are many instances of so-called "quantitative" social science that just define variables *ad hoc*, i.e., without establishing an empirically grounded measurement model to begin with. See Michell (1999) for an in depth discussion of the problems arising from such pseudo-quantitative science.

and the same empirical structure. Nevertheless, the *kinds* of representations (or models) they produce are substantially different. The qualitative strategy emphasizes the meaning of concepts, whereas the quantitative strategy emphasizes the respective topology. Consequently, there is no straightforward way to "translate" a qualitative model into a quantitative model. For example, using the abstract description of a type to inspire a collection of test items that are then scaled by a psychometric model, would not acknowledge the structural difference between qualitative and quantitative modeling approaches. Such a procedure would imply to abandon the typology in favor of a quantitative model, rather than to incorporate the strengths of both kinds of models.

As an alternative, we suggest to translate between qualitative and quantitative models by tracing them back to the empirical structures they are meant to represent. If an empirical structure allows for both, a qualitative and a quantitative representation, the common empirical grounding of the representations ensures that they can be combined in a joint representation where qualitative types are *embedded* in quantitative dimensions.

The key to integration in mixed methods research is thus to approach the data from a qualitative and a quantitative perspective simultaneously. Whereas the qualitative perspective emphasizes the similarity between members of one type and dissimilarity between members of different types, the quantitative structure emphasizes the gradual transitions between different values on a quantitative continuum. However, any sorting by similarity implicitly presumes that there must be *something* with respect to which the objects differ. And since objects can be more or less similar, this something has at least some properties of a quantitative dimension. On the other hand, whenever there are gradual transitions between objects on a quantitative continuum, it is possible to identify some objects that are more alike with respect to this dimension than others. Therefore, any quantitative dimension allows for grouping of objects by similarity.

It is easy to see the connection between qualitative and quantitative representations when objects only differ with respect to a single criterion. For example, in the context of developmental psychology, we may represent individual change over the life span by means *developmental stages*, as proposed by Piaget (1952). Although the concept of a developmental stage is clearly qualitative, it is inherently linked to the concept of *cognitive ability*, which is conceived as a quantitative continuum. Thus, although there may be qualitatively different developmental stages, these stages can be located on a quantitative dimension.

The connection between quality and quantity is somehow less obvious if the empirical structure is more complex. For example, a qualitative reconstruction of teachers' beliefs will most certainly consider several criteria of similarity simultaneously. By definition, beliefs are complex conglomerates of attitudes, thoughts and behavioral dispositions. As a result, the dimensional structure that implicitly underlies a qualitative typology of teachers' beliefs is obscured by the complexity of the field. A trained qualitative researcher may well identify relevant similarities and dissimilarities between the cases. However, it is difficult to construct an underlying quantitative structure from the typology alone.

The above analysis suggests that a special methodology is needed to identify implied quantitative dimensions underlying qualitative typologies. In the following sections, we shall provide such a methodology. Since our approach requires a profound understanding of both, qualitative and quantitative research strategies, we will first elaborate on the specifics of qualitative type formation [with an emphasis on reconstructing "pure types" as characterized by Weber (1904)] and of similarity-based quantitative models (particularly cluster analysis and linear discriminant analysis). We will then outline a general strategy for the embedding of qualitative types in quantitative dimensions.

3. Constructing qualitative types

As outlined above, qualitative research is mainly concerned with the construction of case-based models. Case-based models abstract from singular observations to construct a more general descriptive scheme for the objects under study. For example, if the objects under study are work teams, each team constitutes a singular case. However, each case may also be interpreted as a specific instance of a more general, abstract *type*, which is abductively constructed from qualitative categories that build on comparisons between and within cases. In general, the criteria for these comparisons are not known *a priori*, but emerge from an iterative process of constructing and revising categories (Peirce, 1998; Schurz, 2008).

There are various methodological approaches to the construction of abstract typologies from singular cases (Kluge, 2000). In this article, we focus on the *documentary method*, which analyses qualitative data with regard to the way *how* people talk about certain topics, rather than *what* they say. The rationale behind this shift of focus is the observation that sometimes people's verbal statements seem to contradict their actions. For example, when asked about sustainable behavior, a person may state that the environment is extremely important to her. Nevertheless, she may still fail to implement her stated attitudes in her actions (e.g., taking a hot bath every day instead of a shower or traveling by plane rather than by train). The idea behind the documentary method is that any disparity between what people say and what people do will become manifest in the way people talk about a topic. These different modes of dealing with a topic in a conversation are then used to reconstruct general *patterns of orientation* (Bohnsack, 2010). For example, a person who reports a positive attitude toward sustainable behavior may deal with the topic by emphasizing the political dimension of sustainability and thereby downplay the role of the individual. A different mode of dealing with the topic would be to point toward other, supposedly more important or more urgent, problems such as poverty or war. Both modes point toward different patterns of orientation ("questioning responsibility" vs. "questioning relevance"), which both reveal that the stated positive attitudes most likely differ from actual behavior. The documentary method aims to identify such patterns of orientation to account for the often-observed mismatch between verbal statements and actual behavior and to infer what may be the true motivating forces of peoples' behavior.

In general, the patterns of orientation in a specific context all deal with a common theme – the so-called *tertium comparationis*. In the documentary method, the tertium comparationis provides the interpretative framework for all consecutive analyses. Within this framework, the cases are interpreted as specific realizations of qualitatively different ways of dealing with the tertium comparationis, i.e., different patterns of orientation. For example, "questioning responsibility" and "questioning relevance" are two qualitatively different ways to deal with the common theme of "rationalization of unsustainable behavior." The patterns of orientation are then further condensed into a collection of *pure types*, which together form a qualitative typology that intends to capture all cases within a common interpretative framework. A pure type is not just a descriptive category of what has been observed empirically. It is a theoretical abstraction that transcends the singular cases to form an idealized concept that is reflected in the singular cases but cannot be reduced to them (Weber, 1904). For example, in the context of sustainable behavior, pures type like the "ignorant hedonist" or the "cynical fatalist" may be characterized in such a way that they corresponds to only few (if any) actually observed cases, and yet capture an essential qualitative mode of dealing with the topic of sustainability.

The iterative process of between-cases and within-cases comparisons makes it possible to explicate the pure types in terms of the essential features by which they differ – the *horizons of comparison* that emerge together with the typology (Bohnsack, 2010). These horizons of comparison provide the means to distinguish between the individual cases with respect to the qualities captured by the typology. Like the tertium comparationis and the types themselves, the horizons of comparison are not known *a priori* but are the result of an iterative interpretative process of comparison. All types show different ways of dealing with the tertium comparationis. The horizons of comparison identify the essential properties by which the types differ. For example, the type "ignorant hedonist" may differ from the type "cynical fatalist" with respect to various horizons of comparison, such as the amount of self-efficacy or the amount of social orientation.

The documentary method builds on extensive comparisons within and between cases. These comparisons ensure that the theoretical constructions of the researcher are actually grounded in the empirical material, and that other researchers can retrace their interpretation of the data. However, due to theoretical samplings strategies (Glaser and Strauss, 1979), the documentary method only works with a small or intermediate number of cases, which in turn limits its empirical scope (Bohnsack, 2010). Furthermore, although the documentary method provides a highly systematic rationale for qualitative data analysis, it is not guaranteed that different researchers would arrive at the same results in a specific context. Working in research groups and validating the individually obtained interpretations and theoretical constructions against the critical view of the other members of the research group helps to deal with this problem. However, the results of a Documentary analysis will never be independent of the conducting researchers (i.e., *objective*).

The result of a qualitative analysis based on the documentary method is a theoretically rich typology consisting of an idealized description of the ways that individual cases deal with a certain theme. Since all theoretical concepts emerge from the data by means of constant empirical comparisons, the documentary method is especially useful as a method of theory construction from observations by means of inductive and abductive reasoning.

4. Constructing quantitative dimensions

Whereas qualitative research focuses on case-based models that abstract from singular observations to idealized typologies, quantitative research builds on variable-based models that emphasize the distances of case representations on the dimensions of geometric spaces. However, since any geometric space allows for the calculation of distances between arbitrarily positioned objects, it is always possible to compare objects with respect to their geometric representation and sort them based on their similarities or dissimilarities. In fact, there are various quantitative methods that start with the representation of objects in a geometric space to group them according to their distance with regard to this representation. These methods are subsumed under the label *cluster analysis* (Everitt, 1974). Cluster analysis provides a variety of algorithms to extract groups from the distances of objects in a geometric space. The groups are called clusters and are constructed such that objects within the same group are similar and objects that belong to different groups are dissimilar. The similarity or dissimilarity is measured by a *distance-metric*. A distance metric is a single number that is constructed from the relative positions of two points in a geometric space. Depending on the context, different distance metrics may be appropriate. For example, the distance between two trees on an open field may be measured by their Euclidean distance (i.e., the shortest straight line connecting the two trees). On the other hand, the distance between two houses in downtown Manhattan may be better captured by the city-block metric (i.e., the shortest path a car can take from one block to the other). In the context of quantitative social science, the objects that are compared with regard to their distances are usually individuals. For example, two individuals may be more similar to one another than a third one with regard to their answers on a numeric rating scale in a questionnaire. The squared differences between the individuals' answers to all questionnaire items may then be used to calculate the Euclidean distance of these individuals in an abstract variable space that is spanned by the questionnaire items. Based on the distances between all cases, the individuals are then grouped into homogeneous clusters.

The results of a cluster analysis are actually very similar to the kind of representation produced by qualitative type formation. Nevertheless, the method of construction is completely different. Whereas the qualitative strategy explicates the criterion of similarity by means of extensive comparisons within and between cases, the quantitative strategy starts with a set of variables as the basis of a geometric space and constructs the clusters afterwards. In fact, the result of the analysis is completely determined by the chosen variables, the used distance metric, and the clustering algorithm. Consequently, there is no room for interpretation with regard to the meaning of the clusters or their essential properties. In other words, the dimensions of the geometric space used in cluster analysis are chosen *before* the analysis is performed and are thus not guaranteed to capture the essential properties by which the objects under study differ. Sometimes, the choice of variables is theoretically informed. In other cases, however, variables are chosen without an explicit theoretical frame of reference. Consequently, cluster analysis is less sensitive to new theoretical discoveries than comparative analysis, because the criteria of similarity are specified before the clusters are being constructed and may thus be arbitrary with regard to the essential properties of the clusters. Hence, in contrast to comparative analyses, cluster analysis does not produce pure types in the above sense. It may provide a collection of classes of objects that can be observed empirically, but are not necessarily theoretically meaningful.

Despite these apparent shortcomings, cluster analysis comes with some strong advantages. First, it is mathematically tractable, i.e., both, the clusters and the dimensions of the geometric space, are explicit mathematical objects. They can be precisely defined and communicated in an unambiguous way. Second, the automated algorithms used in cluster analysis ensure that the complete analysis is reproducible once the variables and the distance metric are given. Finally, it is possible to perform cluster analysis with an arbitrary number of objects and an arbitrary number of variables. Therefore, given a sufficient data basis, the results of cluster analysis more easily generalize to larger populations than the results of qualitative type formation.

5. Embedding typologies in geometric spaces

Given the strengths and weaknesses of the qualitative and quantitative approaches outlined above, it would be of great benefit to merge both strategies into an integrated research strategy. Such an integrated approach would aim to construct theoretically rich typologies alongside an explicit geometrical representation of their horizons of comparison as the dimensions of an abstract geometrical space. It is therefore important to not only perform both types of analysis separately, but to ensure that all analogous concepts are continuously translated and back-translated between the different methodological approaches.

In the following, we will outline how such an integrative strategy can be realized. We use techniques borrowed from qualitative reconstructive research such as comparative analyses and the construction of pure types, as well as quantitative modeling techniques like k-means clustering and linear discriminant analysis. The general rationale is to start with an exploratory strategy using a qualitative reconstructive approach, followed by modeling techniques that aim at finding a quantitative representation of the qualitatively discovered pure types that are embedded in a theoretically meaningful geometric space.

5.1. Construct theoretically meaningful types

The first step of our approach exploits qualitative reconstructive techniques of type formation. The aim at this stage is to generate a tentative typology that is both, theoretically rich and empirically grounded. The results do not yet provide a formal, let alone mathematical, description of the empirical structure under study. Nor are they intended to generalize to larger populations. The main purpose is to identify different qualia and to explore their essential features by comparing them with respect to different characteristics. The result should be a collection of pure types that can be distinguished by means of theoretically meaningful horizons of comparison.

The primary techniques used at this early stage of the research are comparative analyses. From a representational perspective, comparative analyses consist in identifying relevant empirical relations between the objects under study. Mathematically speaking, a relation is nothing but a subset of ordered tuples of objects. In the simplest case, the researchers will compare every case to each of the remaining cases and judge them to be either similar enough to be grouped together or not. More complex comparisons may include more than two cases at a time. For example, two cases may be similar to each other when compared to a third case, but not when compared to yet another case, because, in the context of additional cases, the criteria by which one compares the cases may change. It is also possible to judge cases as being similar with respect to one characteristic, but dissimilar with respect to another characteristic. Regardless of the specific comparison procedures, comparative analyses result in a more or less complex relational structure based on the judgments of the researchers.

Since the comparative judgments in a qualitative study are based on interpretative acts, rather than formalized procedures, the resulting structure is, to some degree, subjective. Nevertheless, it is of course possible to assess the degree of consensus between different researchers and adjust the judgments such that they are intelligible across individuals, as it is routinely done in qualitative interpretation groups. Furthermore, comparative analyses naturally imply that researchers reflect on the kind of comparisons they perform as they proceed to analyze the data. Thereby, theoretically fruitful comparisons are eventually identified whereas less useful comparisons are abandoned. This selective component of comparative analyses eventually leads to an abstract representation of the empirical structure by means of an emergent typology, with the types being idealized contrasts (pure types) with regard to theoretically meaningful characteristics (the horizons of comparison).

5.2. Specify initial set of variables

The second step of the analysis takes the pure types and the horizons of comparison as a starting point. The aim at this stage is to identify an initial set of variables that are potentially meaningful with respect to the theoretical typology constructed in the first step.

The main technique used in this step is to extract the most relevant and most specific characteristics of the pure types constructed in step one and to transform them into a questionnaire. For some characteristics, this may be straightforward. Other, more abstract ones, may require more attention. For example, an abstract concept like "relativism" is way too vague to include it in a questionnaire. Consequently, researchers have to partly de-construct the abstract characteristics of the pure types to arrive at a set of unambiguous characteristics that can be transformed into questionnaire items (e.g., "What is true for one person may be untrue for another person."). Note that we are not dealing with some kind of latent variable here–"relativism" is not something unobserved underlying the concrete characteristics we want to include in the questionnaire. It is a theoretical abstraction that results directly from comparative analyses (Buntins et al., 2016). Therefore, the relation between the questionnaire items and the abstract construct is not one of cause and effect, as implied by latent variable models like classical test theory or item response theory. It is a logical relation that depends strictly on the way the researchers use the abstract theoretical vocabulary that emerged from comparative analyses (*cf.* Buntins et al., 2017; Borgstede, 2019; Leising and Borgstede, 2019; Borgstede and Eggert, 2022). Thus, we can only decide whether a question belongs in our questionnaire on theoretical grounds. Since we are interested in quantitative comparisons in the consecutive steps, it is reasonable to use some kind of numeric answer type (e.g., a Likert scale). However, note that the use of numerical scales does not necessarily imply psychological measurement of an unobserved variable (Michell, 1999).

For the same reasons, standard psychometric criteria like convergent and divergent validity or internal consistency are largely irrelevant with regard to the questionnaire resulting from the characterization of the pure types. In fact, the only relevant criterion for the questionnaire is that the translation between the abstract theoretical constructs from the qualitative analysis to concrete statements in everyday language is successful (*cf.* Buntins et al., 2017). Like the qualitative analysis itself, the construction of a questionnaire that is valid in this respect is subject to interpretation and thus requires a corrective in the form of critical discussion between researchers.

5.3. Ensure generalizability

Step three consists in assessing the theoretically derived characteristics from step one in a large sample using the questionnaire developed in step two. The aim at this stage is to ensure generalizability of the typology that was tentatively developed in step one.

The main issue to be addressed when distributing the questionnaire is the intended scope of the theory. Since our tentative typology was constructed using qualitative methods, its empirical grounding will most certainly be limited to a rather small number of cases. These cases are not a random sample but the result of a purposive sampling strategy called theoretical sampling. Theoretical sampling means that cases are not selected before the data analysis, but as the theory evolves. Such a strategy implies that data collection, data analysis and theory development are parallel processes that influence one another (Glaser and Strauss, 1979). The intended scope of the theory is thus just as much an emergent property of comparative analyses as the developed typology and the horizons of comparison. Consequently, before administering the questionnaire on a large scale, it is important to analyze the results of the qualitative analysis with regard to the cases that were sampled. Most importantly, one has to identify the common characteristics of the cases (as opposed to the previous step, which consisted in identifying the differences between cases). These characteristics serve as a first characterization of the population one wishes to describe. In a way, they provide the boundary conditions of the theory that is being developed.

Of course, there are many characteristics that are too general to be useful as boundary conditions. For example, the majority of human subjects have two legs, two eyes, a nose etc. Unless one of these characteristics is theoretically relevant (e.g., when the typology aims to describe various forms of discrimination against people with disabilities), they are rather uninformative and thus useless as boundary conditions. On the other hand, when the common characteristics of the cases are too specific, the resulting population may consist of very few cases, and may even be restricted to only those cases that have actually been sampled. Therefore, an intermediate level of abstraction is required to produce a workable best guess about the scope of the theory.

Based on these theoretical considerations, the target population for the questionnaire assessment can eventually be specified. Once the population is known, the best strategy is to adopt a random sampling strategy and to sample as many cases as possible. Note that *a priori* considerations about statistical power are not applicable, since we are still in an exploratory stage. Consequently, the aim of the questionnaire study is not to test statistical hypotheses, but to provide a representative data source for a quantitative embedding of the qualitative typology constructed earlier.

5.4. Formalize initial typology

We now enter the first stage of quantitative modeling. Step four consists in applying mathematical algorithms to the data obtained in the previous step that produce a geometric representation of the characteristics identified in step two alongside a set of empirically derived clusters. The aim of step four is to produce an initial quantitative approximation to the pure types constructed in step one.

As outlined in section 4, we propose to use statistical clustering methods to identify homogeneous groups of objects based on the variables constructed from the numerical answers to the questionnaire administered in step three. Since we already have a tentative theory in the form of pure types, we will use an algorithm that produces a pre-determined number of clusters. The corresponding method is called k-means cluster analysis (MacQueen, 1967).

The basic idea behind k-means clustering is that the variables assessed in the questionnaire are interpreted as the dimensions of an abstract geometric space. The notion of "space" is similar to the standard use of the word for the three-dimensional space we use to describe the position and movement of physical objects. However, an abstract space constructed from the numerical answers in a questionnaire can have an arbitrary number of dimensions, each for every question. Moreover, this abstract variable space does not correspond to a real physical object. It is barely more than a quantitative representation of an implicit distance structure. In other words, the abstract geometric space only serves the purpose of providing a distance metric between objects allocated in the space. This distance metric is then used to group objects according to their similarity, where "similar" means that the distance is small and "dissimilar" means that the distance is large.

The k-means clustering algorithm searches for a partition of the objects into the specified number of clusters such that the average deviation of the objects from the center of their assigned cluster is as small as possible. The standard algorithm starts with an arbitrary initial partition of objects into k clusters and calculates the mean values of the objects in each cluster (hence the name k-means clustering). The partition is then updated by re-assigning each object to the cluster that is closest with regard to the distance metric. After the re-assignment, the cluster means are re-calculated using the new members of the clusters. The procedure is repeated until the clusters do no longer change (Lloyd, 1982).

The result of this k-means cluster analysis is a mathematically unambiguous partitioning of the cases that were sampled in step three with regard to the characteristics identified in step two that were themselves derived from the pure types constructed in step one. We have thus a first quantitative approximation to the pure types that made up our tentative theory.

5.5. Formalize horizons of comparison

The steps one to four were concerned with a first translation between qualitative types and quantitative dimensions. However, the geometric space constructed in the previous steps is not yet a faithful formalization of the horizons of comparison that differentiate between the pure types. Neither are the clusters obtained in step four formal counterparts to the pure types themselves. Due to the data-driven approach of the k-means clustering algorithm, the clusters correspond to *real types*, rather than pure types. They do not abstract from the raw data on a theoretical level but rather average over the objects that belong to the same cluster. Therefore, in the last two steps of the analysis, the quantitative representation constructed so far will be transformed such that the clusters will be pure types (in the sense that prototypical characteristics are emphasized), and the corresponding geometric space is spanned by a set of abstract dimensions that differentiate maximally between these pure types.

The method employed at this stage is linear discriminant analysis (Klecka, 1980). Linear discriminant analysis is a statistical method that transforms one geometric space into another geometric space, such that the newly constructed dimensions differentiate maximally between groups of objects. In the simplest case, one starts with a collection of objects that are divided into two groups (say, group A and group B). Given a set of quantitative variables that describe the individual objects, the method now calculates a weighted sum over these variables to obtain a so-called discriminant variable. The weights in the summation are chosen such that members of group A have, on average, low values on the discriminant variable, and members of group B have, on average, high values on the discriminant variable. The weights are adjusted until the discriminant variable differentiates maximally between the two groups. Linear discriminant analysis can also be applied if the objects are divided into more than two groups. When the groups only differ along one continuum, the method will still only yield one discriminant variable. Else, the method will produce a set of several discriminant variables that form a multidimensional discriminant space.

When applied to the results of a cluster analysis, linear discriminant analysis yields an alternative, more abstract geometrical embedding for the clusters. To emphasize the differences between the clusters, the discriminant space is routinely projected onto a space of lower dimensionality (usually, two dimensions are sufficient). In contrast to the initial set of variables used to construct the clusters (which were merely a best guess about the relevant horizons of comparison), the discriminant space is constructed such that it provides the most parsimonious and efficient way to distinguish the clusters from one another. Therefore, the result of step five is a new geometric embedding for the clusters obtained in step four, that corresponds to a set of abstract horizons of comparison by which the clusters can be distinguished.

5.6. Formalize pure typology

The final step of the analysis consists in a formal method to construct pure types that are embedded in a quantitative geometric space consisting of a small set of maximally differentiating dimensions. The corresponding geometric space has already been constructed in the previous step. The aim of the last stage of analysis is to use this abstract geometric space to update the initial clusters constructed in step four.

The linear discriminant space captures the abstract dimensions that differentiate maximally between the initial clusters. To construct new, idealized, clusters (that correspond to the pure types tentatively proposed in step one) we apply a second k-means cluster analysis—only this time we use the discriminant variables (instead of the initial variables) to define the distance metric. Like before, we use the same number of clusters as in our initial typology. Since the discriminant variables are determined from the initial variables, which in turn were constructed from the description of the qualitative criteria of similarity in step two, both, the geometric space and the clusters constructed in this last step, are still grounded in the empirical structure we wish to describe. However, in contrast to a simple cluster analysis (that yielded real types), we now have clusters in a discriminant space.

Since the discriminant space captures the abstract characteristics by which the clusters differ, these new clusters are maximally different with respect to the initial grouping. In other words, those characteristics that strongly differentiate between clusters are weighed more strongly than those that differentiate poorly. Consequently, the updated clusters are idealizations of the original, empirically derived clusters. These idealized clusters capture the relevant horizons of comparison by emphasizing those aspects in the data that are specific to the clusters. Therefore, the idealized clusters correspond to pure types as characterized in section 3.

The result of the last step of analysis is a formal reconstruction of a tentative qualitative typology based on a large sample of cases. This formal reconstruction captures the implicit quantitative structure underlying the act of grouping objects by similarity, as well as the qualitative aspects of the empirical structure that is characterized by pure types.

6. Exemplary application: Teachers' epistemological beliefs

In the previous section, we presented an integrated approach to embed qualitative pure types in a quantitative geometric space to form an overarching model of the underlying empirical structure. Our approach builds on a representational integration of qualitative and quantitative modeling strategies as outlined by Borgstede and Scholz (2021).

We outlined how qualitative typologies that emerge from comparative analyses can be interpreted as an attempt to construe abstract relational structures that are grounded in empirical observations. Building on this abstract conception of qualitative typologies, we proposed that the horizons of comparison from such a typology be translated into a collection of questionnaire items which are then used as an initial variable space for a k-means cluster analysis. Using linear discriminant analysis to construct an abstract geometric space of maximally discriminating dimensions, we then proposed to revise the initial cluster solution based on these abstract dimensions. This procedure yields a formal representation of pure types within a quantitative space consisting of abstract dimensions.

In this section, we will illustrate the feasibility of our method by means of an exemplary application. The application is concerned with the reconstruction of teachers' epistemological beliefs and builds on a published qualitative reconstructive analysis (Rau, 2020, 2021). We will start with a brief review of the theoretical background of the study and the results of the qualitative type formation. The main part of the example focuses on the concrete procedure of formalizing the typology proposed in Rau (2020) using the integration approach outlined above. All statistical analyses were preformed using R version 4.0.3 (R Core Team, 2020) with the additional packages janitor (Firke, 2021), MASS (Venables and Ripley, 2002) and factoextra (Kassambara and Mundt, 2020).

6.1. Background

Rau (2020) conducted a qualitative reconstructive study using the documentary method (*cf.* section 3). The study examined epistemological beliefs of teachers who teach a humanities subject. The aim of the study was to describe how teachers generate knowledge about cultural artifacts in the classroom. The study focused on the following questions: (a) How do teachers deal with cultural artifacts such as poems or images in their teaching practices? (b) How do teachers interact with their students to flesh out the meanings of these cultural artifacts? The data were collected by group discussions ($N = 19$). Cases were selected according to interviewees' characteristics (e.g., level of education, expert or novice in the teaching profession, subject studied) following a theoretical sampling strategy (Rau, 2020, 2021).

The documentary method works by abstracting findings and finding a common theme that is common to all cases: the tertium comparationis. The tertium comparationis identified in the study refers to the way that teachers justify their understanding of cultural artifacts and was thus coined *justification*. Systematic comparisons within and between cases revealed three basic ways of dealing with justification, each of which constitutes a pure type in the sense of Weber (1904). The three types were: (1) Contingency stops, (2) Orientation to application (3) Appreciation of pluralism. These three types comprise the basic constituents of teachers' epistemological beliefs. They indicate how teachers' epistemological beliefs may guide instructional action in the humanities and how justification of different meanings is ensured. Teachers of type 1 expect pupils to interpret cultural artifacts within their historical context of origin and to elaborate on authorial intention by choosing a deductive method. Pupils have to learn epochal knowledge and are not allowed to bring their own meanings for the cultural artifact into the discourse. Teachers of type 2 ask, what personal meaning cultural artifacts have for the individual pupils. The learners' ability to articulate their personal meaning of the cultural artifact is a characteristic of justification for teachers. These teachers are eager to learn about different meanings that pupils give to the same cultural artifact. Teachers of type 3 choose different (theoretical) perspectives to look at cultural artifacts. For example, they may interpret cultural artifacts with a feminist reading. The empirical material shows that these teachers engage in a discursive classroom conversation with their pupils to agree on an interpretation of the cultural artifact. It is important to these teachers that the pupils adopt a critical attitude and learn to reflect on their own about the meaning of cultural artifacts. Justification is based on intersubjective validation between the teacher and the pupils.

The within-cases and between-cases comparisons that generated the tertium comparationis and the pure types also revealed the main characteristics by which the three types may be distinguished. These characteristics constitute the essential horizons of comparison by which the identified, abductively formed, qualia differ. These horizons of comparison were: (a) Genesis of meaning: Which epistemological approaches do teachers choose? (b) Certainty and limits of generated knowledge: What do teachers expect from justification? (c) Characteristics of the cultural artifacts: What are the ontological characteristics of cultural artifacts in the humanities? Do teachers perceive cultural artifacts as ambiguous and/or unambiguous? (d) Relating teachers to pupils: How do teachers include pupils in the genesis of meaning and knowledge? To what extent do teachers allow pupils to discuss their own attributions of meaning in class? (e) Aims of teaching in the humanities: What are the aims of teachers' teaching in relation to epistemic learning of the pupils? The three pure types, as well as the five horizons of comparison are summarized in Table 1. The table also shows how the different types can be distinguished with respect to these horizons of comparison.

TABLE 1 Qualitative typology obtained from documentary method.

	Type 1: *Contingency stops*	Type 2: *Orientation to application*	Type 3: *Appreciation of pluralism*
Horizon of comparison a: *Genesis of meaning*	Historicizing; monoperspectival	Relativistic	Multi-perspectival
Horizon of comparison b: *Certainty and limits of generated knowledge*	Deductive method	Extension of decoding performance	Advance of knowledge; intersubjective validation
Horizon of comparison c: *Characteristics of the cultural objectivation*	Unique; unambiguous	Ambiguous	Ambiguous
Horizon of comparison d: *Relating teachers to pupils*	Exclusion of students from the discourse	Supplementing teacher knowledge with pupils' views	Conversation about decoding
Horizon of comparison e: *Aims of teaching in the humanities*	Imparting of epochal knowledge	Search for further knowledge; surplus value for pupils	Practicing a critical attitude

6.2. Data

The qualitative typology put forward in Rau (2020) was used to construct a questionnaire. The questionnaire contained 43 items that were formulated such that they capture the five horizons of comparison that differentiate between the three pure types. Since these horizons of comparison were formulated on a high level of abstraction in the original study, they were first concretized and translated into everyday language, such that respondents were able to understand them correctly. For example, the first pure type ("contingency stop") can be characterized with respect to the first horizon of comparison ("genesis of meaning") by the fact that teachers aim to convey the meaning that the originator (supposedly) ascribed to the cultural artifact to their students. The corresponding item generated to capture this aspect of the horizon of comparison was: "When pupils interpret pieces of music, literature or art, it is important that they carve out the author's intention." To ensure item comprehensibility and content validity, the initial item set was presented to two teachers in the field of humanities and two experts in the field of teachers' epistemological beliefs. Critical feedback from these expert judgments was incorporated in a revised item set, which was then used for a first empirical study.

The sample consisted of 153 undergraduate students from a Bavarian University with a focus on teachers' education and humanities. The students were recruited in university seminars and lectures. 83.1% of the respondents identified as female, 15.5% as male, and 1.4% as diverse. The median age of the participants was 19 years with an inter quartile range of 4 years. 13.3% of the sample reported a migration background. 54.9% of the participants were student teachers for primary schools, 22.5% were student teachers for vocational schools, 18.3% were student teachers for high schools, and 4.2% reported a different type of school. 19.6% of the students reported at least some kind of teaching experience, although most students (80.5%) were not teaching at the time of the study.

The statistical analyses followed the steps described in section 5. First, a k-means cluster analysis was conducted using the numerical answers to the questionnaire items as a variable space and the Euclidean distance as a distance metric. Since the qualitative analysis identified three pure types, a three-cluster model was fit to the data. All data were standardized within individuals before the analysis. Cases with missing values were excluded. The so-obtained clusters were then used as grouping variables in a linear discriminant analysis. Finally, a second k-means cluster analysis was conducted with the discriminant variables as a variable space.

6.3 Results and discussion

Figure 1 presents the results of the initial k-means cluster analysis that was conducted using the questionnaire items as a variable space. Since the actual variable space contains 43 dimensions (one for each questionnaire item), the data was projected into a two-dimensional space to enable a graphical depiction of the clusters. The two dimensions in the figure were chosen as a reference coordinate system such that they bind the maximum amount of variance in the data.[3]

The three-cluster solution shows that there are two clusters (cluster 1 and cluster 3) that are completely separate. These two clusters show no systematic differences in the y-axis but are clearly distinct with regard to the x-axis. Cluster 2 takes an intermediate position on the x-axis and shows higher average values on the y-axis, indicating that it differs from the other two clusters with regard to a distinct dimension. However, cluster 2 has considerable overlap with the other two clusters, especially with cluster 3. Therefore, it remains open at this stage of analysis, whether it actually captures a different qualitative aspect of the data.

The three clusters from the original variable space were then used as grouping variables in a linear discriminant analysis. Two discriminant variables were constructed such that they are independent of one another and differentiate optimally between the three clusters. A second k-means cluster analysis with three clusters was conducted using the discriminant variables as a variable space (see Figure 2). Since the principle components coincide with the discriminant variables, the x- and y-axes of Figure 2 can now be interpreted as the two dimensions that differentiate maximally between the clusters and bind a maximal amount of overall variance of the data.

Like in the initial cluster solution, clusters 1 and 3 differ mainly in one dimension, which is indicated by the x-axis in Figure 2. However, in contrast to the initial solution, cluster 1 now has a higher overall within-cluster variance than clusters 3 and 2, indicating that it is less homogeneous than the other two. Like before, cluster 2 takes an intermediate position on the x-axis and higher average values on the y-axis, indicating that it mainly differs from the other two clusters with regard to a different dimension.

In contrast to the initial solution, the updated clusters do no longer overlap. This is a direct result of the newly constructed variable space. Since the discriminant variables are specified such that they maximally differentiate between the clusters, existing differences between the original clusters are emphasized because those questionnaire items that differentiate more receive higher weights when calculating the linear discriminant variables (which are, in fact, just weighted sums of the original variables). The resulting clusters are thus idealizations of the original clusters. Just like the pure types in a qualitative typology, they are an abstract representation of the observed qualia rather than a purely descriptive summary as it is given by real types. Note that the data has not been changed to obtain this idealized cluster solution. The individual item answers are the same as before. The only difference to the initial solution is that the coordinate system has been changed by means of a linear transformation (i.e., a weighted summation), such that the coordinates differentiate maximally between the clusters. Thus, the qualitative differences that were only partly visible in the initial solution are revealed, alongside an abstract coordinate system that represents the primary axes by which the clusters differ.

The last step of the analysis aimed to identify the semantic content of the three idealized clusters. To get an impression about which questionnaire items contribute most to the two discriminant variables, the corresponding weighting factors were inspected. Comparing the items with the highest and lowest relative weights for each of the discriminant variables, it turned out that the first dimension (the x-axis in Figure 2) corresponds to the distinction between a pluralistic view (e.g., "Pupils should interpret pieces of music, literature or art from different perspectives.") and a dogmatic view on the meaning of cultural artifacts (e.g., "It is important that pupils interpret pieces of music, literature or art in line with established views from the scientific community."). The second dimension (y-axis in Figure 2) corresponds to the distinction between an orientation toward the past

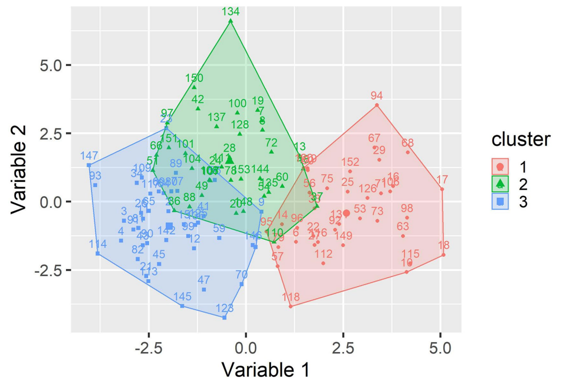


FIGURE 1
Initial cluster solution projected into two dimensions.

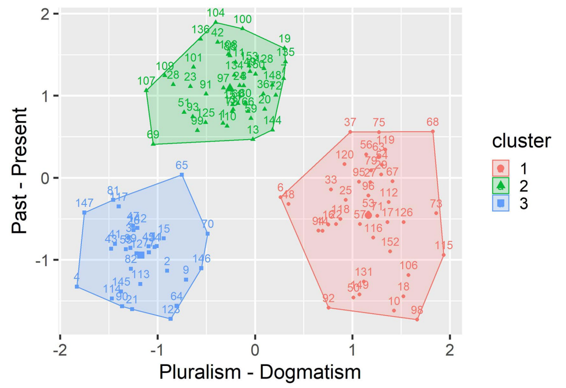


FIGURE 2
Idealized cluster solution obtained from linear discriminant variables.

3 Formally, these dimensions correspond to the first two variables obtained from a principle component analysis (*cf.* Pearson, 1901).

(e.g., "I usually incorporate a history-dependent view on scientific knowledge into my classes.") and an orientation toward the present with regard to the meaning of cultural artifacts (e.g., "The current life-world of the pupils affects how they interpret pieces of music, literature or art.").

In light of these results, we can contrast the three idealized clusters as follows. Cluster 1 and 3 primarily differ on the pluralism-dogmatism continuum, with cluster 3 leaning toward pluralism and cluster 1 toward dogmatism. Cluster 2 can be characterized as leaning toward pluralism, as well, although less than cluster 3. The main difference between clusters 2 and 3, however, is not the degree of pluralism but that cluster 2 is characterized by an orientation toward the present, whereas cluster 3 leans toward a past orientation. This tendency toward the present includes the current environment of the students and acknowledges the relevance of the students' own experiences and perspectives on the meaning of music, art and literature.

These characterizations capture essentially the same qualia as the pure types from the original study. The fundamental horizons of comparison were also reproduced almost exactly as in the qualitative analysis. We can thus conclude that the qualitative pure types that were identified by means of comparative analyses can in fact be formalized as idealized clusters in a quantitative geometric space. The resulting representation converges nicely with the qualitative analysis, thereby sharpening the verbal descriptions within a mathematical model that builds on data from a larger sample. The model captures both, the qualitative and the quantitative aspects of teachers' epistemological beliefs, because it builds on a methodologically well-founded integration of qualitative and quantitative research strategies within a representational framework.

7. Conclusion

This article dealt with the question how qualitative and quantitative research strategies can be integrated such that qualitative types and quantitative dimensions are represented within the same overarching model. We argued that a true integration can only be achieved if qualitative and quantitative modeling strategies are viewed in light of a common methodological framework. The representational approach put forward by Borgstede and Scholz (2021) provides such a methodological background. In this paper, we refined the representational approach and applied it to an empirical case study, thereby demonstrating how qualitative and quantitative methods can be merged to produce formal representations that capture both, qualitative and quantitative, aspects in the data and integrate them within a single model.

Our approach transcends the distinction between qualitative and quantitative research by providing a common conceptual framework. Within this framework, it is possible to translate between qualitative and quantitative modeling approaches and to facilitate the simultaneous discovery of both kinds of structures. For example, concepts like "pure types" generally have no meaning in quantitative research. However, from a representational perspective, pure types can be conceived as idealized clusters in an abstract geometric space. Similarly, the concept of a "distance metric" has no meaning in qualitative research. However, in light of the representational view, a distance metric is just a formalized version of the criterion of similarity or dissimilarity used to compare the objects under study.

In this article, we focused on the question how qualitative typologies can be embedded in a quantitative geometric space. However, our approach provides a far more general rationale for the construction of new research designs. The essential point is to realize that qualitative and quantitative models are just different kinds of abstract relational structures and that they both attempt to represent empirical relational structures. Following this rationale, qualitative comparisons within and between cases may be considered as the empirical basis for various quantitative scaling techniques. Similarly, quantitative representations may be exploited to extract qualitative distinctions within and between cases. For example, the similarity judgments of qualitative researchers may be used as primary data for the construction of an abstract feature space by means of multidimensional scaling (Borg et al., 1997). The abstract feature space can then be compared to the horizons of comparison derived from comparative analyses. Other applications might include psychometric models with qualitative components (like multi-group Rasch models or latent class analysis, *cf.* von Davier and Carstensen, 2007), the embedding of qualitative contingencies in an abstract variable space by means of correspondence analysis (Hirschfeld, 1935), or the application of mathematical algorithms to identify specific similarities between cases by means of formal concept analysis (Ganter and Wille, 1999).

The integration strategy outlined above requires profound knowledge of both, qualitative and quantitative, modeling techniques. In particular, the translation between different kinds of models depends on an abstract understanding of the empirical and theoretical structures involved in the analysis. Such an abstract understanding requires a level of formalization that is rarely achieved in empirical research, let alone in the context of theory building in educational science. Although formal approaches to empirical research and theory formation may be challenging and sometimes seem cumbersome, we think that they are worth the effort. Our research example shows how the results from qualitative and quantitative analyses to the same data may converge using our representational approach. Moreover, the study shows that the proposed strategy of integration enriches the theoretical scope of the quantitative model components and scrutinizes the semantic import of its qualitative aspects.

The representational approach emphasizes the similarities between qualitative and quantitative research strategies and provides a metatheoretical framework to identify relevant differences at the same time. We hope that this overarching perspective will not only find its way into mixed methods research, but also facilitate communication and foster mutual exchange between qualitative and quantitative researchers in general.

Data availability statement

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

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MB conceived the conceptual background, developed the methodology, conducted the statistical analyses, made the visualizations, and wrote the first draft of the manuscript. CR verified the conceptual background and collected the data. All authors contributed to the article and approved the submitted version.

Acknowledgments

We thank Jana Costa for her help with the formulation of the questionnaire items.

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

Bazeley, P. (2016). Mixed or merged? Integration as the real challenge for mixed methods. *Qual. Res. Organ. Manage.* 11, 189–194. doi: 10.1108/QROM-04-2016-1373

Bergman, M. M. (2011). The good, the bad, and the ugly in mixed methods research and design. *J. Mixed Methods Res.* 5, 271–275. doi: 10.1177/1558689811433236

Bohnsack, R. (2010). "Documentary method and group discussions" in *Qualitative analysis and documentary method in international educational research*. eds. R. Bohnsack, N. Pfaff and W. Weller (Opladen: Barbara Budrich), 99–124.

Borg, I., and Groenen, P. J. F. (Eds.) (1997). "Springer series in statistics" in *Modern multidimensional scaling: Theory and applications* (Springer)

Borgstede, M. (2019). Zwischen Definition und Empirie. *Vierteljahrsschrift Für Wissenschaftliche Pädagogik* 95, 199–217. doi: 10.30965/25890581-09501018

Borgstede, M., and Eggert, F. (2022). Squaring the circle: from latent variables to theory-based measurement. *Theory Psychol.* 33. doi: 10.1177/09593543221127985

Borgstede, M., and Scholz, M. (2021). Quantitative and qualitative approaches to generalization and replication—A representationalist view. *Front. Psychol.* 12:605191. doi: 10.3389/fpsyg.2021.605191

Buntins, M., Buntins, K., and Eggert, F. (2016). Psychological tests from a (fuzzy-)logical point of view. *Qual. Quant.* 50, 2395–2416. doi: 10.1007/s11135-015-0268-z

Buntins, M., Buntins, K., and Eggert, F. (2017). Clarifying the concept of validity: from measurement to everyday language. *Theory Psychol.* 27, 703–710. doi: 10.1177/0959354317702256

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles: Sage.

Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Los Angeles: Sage.

Everitt, B. (1974). *Cluster analysis* Hoboken: John Wiley & Sons.

Fetters, M. D., and Molina-Azorin, J. F. (2017). The journal of mixed methods research starts a new decade: the mixed methods research integration trilogy and its dimensions. *J. Mixed Methods Res.* 11, 291–307. doi: 10.1177/1558689817714066

Fielding, N. G. (2012). Triangulation and mixed methods designs. *J. Mixed Methods Res.* 6, 124–136. doi: 10.1177/1558689812437101

Firke, S. (2021). Janitor: Simple tools for examining and cleaning dirty data. Available at: https://CRAN.R-project.org/package=janitor

Flick, U. (2014). *The SAGE handbook of qualitative data analysis*. London: Sage.

Freeman, M., deMarrais, K., Preissle, J., Roulston, K., and Pierre, E. A. (2007). Standards of evidence in qualitative research: an incitement to discourse. *Educ. Res.* 36, 25–32. doi: 10.3102/0013189X06298009

Ganter, B., and Wille, R. (1999). *Formal concept analysis*. Berlin Heidelberg: Springer.

Glaser, B. G., and Strauss, A. L. (1979). *The discovery of grounded theory: Strategies for qualitative research (10. Printing)*. New York: Aldine.

Hirschfeld, H. O. (1935). A connection between correlation and contingency. *Math. Proc. Camb. Philos. Soc.* 31, 520–524. doi: 10.1017/S0305004100013517

Kassambara, A., and Mundt, F. (2020). Factoextra: extract and visualize the results of multivariate data analyses. 1–84. Available at: https://CRAN.R-project.org/package=factoextra

Klecka, W. R. (1980). *Discriminant analysis. Quantitative applications in the social sciences series*, vol. *19*. Thousand Oaks: Sage.

Kluge, S. (2000). Empirically grounded construction of types and typologies in qualitative social research. *Forum qualitative Sozialforschung/forum: Qual. Soc. Res.* 1:1. doi: 10.17169/fqs-1.1.1124

Krantz, D. H., Luce, D., Suppes, P., and Tversky, A. (1971). *Foundations of measurement volume I: Additive and polynomial representations* New York, London: Academic Press.

Leising, D., and Borgstede, M. (2019). "Hypothetical constructs" in *Encyclopedia of personality and individual differences*. eds. V. Zeigler-Hill and T. K. Shackelford (Cham: Springer International Publishing), 1–6.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory* 28, 129–137. doi: 10.1109/TIT.1982.1056489

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proc. 5th Berkeley Symp. Math. Stat. Probab. Symposium conducted at the meeting of Univ. Calif.

Michell, J. (1999). *Measurement in psychology* Cambridge: Cambridge University Press.

Moran-Ellis, J., Alexander, V. D., Cronin, A., Dickinson, M., Fielding, J., Sleney, J., et al. (2006). Triangulation and integration: processes, claims and implications. *Qual. Res.* 6, 45–59. doi: 10.1177/1468794106058870

O'Cathain, A., Murphy, E., and Nicholl, J. (2010). Three techniques for integrating data in mixed methods studies. *BMJ* 341:c4587. doi: 10.1136/bmj.c4587

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *London Edinburgh Dublin Philos. Magazine J. Sci.* 2, 559–572. doi: 10.1080/14786440109462720

Peirce, C. S. (1998). "The essential Peirce" in *Selected philosophical writings*. eds. N. Houseret al., vol. *2* (Bloomington: Indiana Univ. Press), 1893–1913.

Piaget, J. (1952). *The origins of intelligence in the child* (M. Cook, Trans.). New York: W. W. Norton & Co

R Core Team. (2020). R: A language and environment for statistical computing. Available at: https://www.R-project.org/

Ragin, C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.

Rasch, G. (1960). *Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut.

Rau, C. (2020). *Kulturtradierung in geisteswissenschaftlichen Fächern. Eine rekonstruktive Studie zu epistemologischen Überzeugungen von Lehrkräften. Teaching culture in the humanities. A reconstructive study of teachers' epistemological beliefs*. Bad Heilbrunn: Verlag Julius Klinkhardt.

Rau, C. (2021). Die Wissensgrundlagen des eigenen Fachs verstehen – empirische Befunde zu den epistemologischen Orientierungen von Lehrkräften geisteswissenschaftlicher Fächer. *Z. Erzieh.* 24, 91–112. doi: 10.1007/s11618-021-00992-y

Rihoux, B., and Ragin, C. (2009). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Thousand Oaks: SAGE Publications, Inc.

Schurz, G. (2008). Patterns of abduction. *Synthese* 164, 201–234. doi: 10.1007/s11229-007-9223-4

Venables, W. N., and Ripley, B. D. (2002). Modern applied statistics with S (fourth). Springer. Available at: https://www.stats.ox.ac.uk/pub/MASS4/

von Davier, M., and Carstensen, C. H. (eds.) (2007). *Multivariate and Mixture Distribution Rasch Models - Extensions and Applications*. New York: Springer.

Weber, M. (1904). "Objectivity in social science and social policy" in *The methodology of the social sciences*. eds. E. A. Shils and H. A. Finch (New York: Free Press)

Yin, R. K. (2006). Mixed methods research: are the methods genuinely integrated or merely parallel. *Res. Sch.* 13, 41–47.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/fpsyg.2023.1015759

Check for updates

OPEN ACCESS

EDITED BY
Antonio P. Gutierrez de Blume,
Georgia Southern University,
United States

REVIEWED BY
Paolo Renna,
University of Basilicata,
Italy
Xie Yu,
Northwestern Polytechnical University,
China

*CORRESPONDENCE
Xiaomeng Sun
✉ sunxiaomeng@shzu.edu.cn

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

RECEIVED 10 August 2022
ACCEPTED 13 January 2023
PUBLISHED 16 February 2023

CITATION
Sun XM (2023) Higher education management in western regions by educational power strategy and positive psychology.
Front. Psychol. 14:1015759.
doi: 10.3389/fpsyg.2023.1015759

Higher education management in western regions by educational power strategy and positive psychology

Xiaomeng Sun*

Normal College, Shihezi University, Shihezi, China

With the deepening of the strategy of strengthening the country through education, the innovation and development of higher education, system reform and teaching innovation in the western region have become the focus of researchers' attention, and the optimization of educational power strategy has always been an important basis for the development of teaching work. On the basis of fuzzy models Takagi and Sugeno (T–S), this paper constructs an educational resource recommendation model based on T–S fuzzy neural network, verifies the feasibility of the model, further combines the educational resource recommendation model with university teaching, and analyzes the application effect. The current situation of educational resources investigation in M College is analyzed. It is found that the full-time teachers' overall academic qualifications are not high, the proportion of young full-time teachers with certain experience is small, and the professional advantages of the school are not obvious. After applying the educational resource recommendation model, the accuracy of educational resource recommendation is obviously improved, and the design is feasible. The educational management mode with positive psychological emotions has a good teaching effect, which can greatly improve teachers' dedication and concentration. Positive psychological emotions can reduce the possibility of intensification of contradictions and the possibility of behavioral opposition. Teaching resource recommendation mode can improve college students' interest in the application of teaching resources to a certain extent, and their application satisfaction is obviously improved. This paper not only provides technical support for the improvement of teaching management resource recommendation mode, but also contributes to the optimization of teaching power strategy.

KEYWORDS

strategy of strengthening the country through education, positive psychology, western region, higher education, educational resources

1. Introduction

With the continuous strengthening of China's economic strength, building an educational power has become a necessary task for China to promote the development of higher education (Dopson et al., 2019). In particular, it plays an important role in the construction of higher education in western China. Due to the limitations of various conditions, the overall level of education in western colleges and universities has no advantages compared with that in the eastern region. Strengthening the construction of higher education in the western region has become an important construction task for implementing the strategy of strengthening the country through education.

In the context of continuous social development, higher education has become an important cornerstone of current social development. College and university education are not only important sources of high-quality talent output but also an important base for scientific and technological innovation. Therefore, strengthening the innovation and development of college and university education is one of the main tasks of current society. However, at present, many colleges still do not pay enough attention to education reform under the premise of building a strong educational country. The traditional school education development model has greatly reduced the quality of education, and it is difficult to achieve the construction of a strong education country (Runhong, 2022). In order to implement the strategy of strengthening the country through education, colleges and universities need to recognize the fundamental tasks of education, innovate training objectives, implement teaching reform, optimize curriculum institutions, improve the level of teaching activities, and strengthen the construction of teaching staff. The construction of higher education in the western region plays an important role in the construction of a powerful country in higher education due to its special orientation (Dong et al., 2019). In recent years, due to the constraints of geographical location, economic development, teaching scientific settings, education mode, and other factors, China's higher education has shown a pattern of "the east region is strong and the west region is weak." It causes a large academic brain drain in the western region, the rigid personnel management system, weak social service ability, and other issues (Fu et al., 2021). Topics such as innovation and the development of higher education in the western region, system reform, teaching innovation, and entrepreneurship education have become the focus and research objects of all walks of life. Carlucci et al. (2018) proposed a framework in the study to analyze students' evaluations of the quality of higher education teaching and to reveal the risks that affect the quality of teaching and the quality of courses requiring continuous improvement. The framework integrates two decision-based approaches: a standardized U control chart and ABC analysis using fuzzy weights. Using student ratings, control charts can identify courses that need to improve teaching quality in the short term. ABC analysis uses fuzzy weights to deal with the fuzziness and uncertainty of students' teaching evaluations and provides a risk map of potential areas for long-term teaching performance improvement. The proposed framework allows for the prioritization of corrective measures needed to respond to student criticism of teaching and curriculum quality (Carlucci et al., 2018). Li et al. (2020) used the comparative analysis method to conduct a vertical evolution analysis and a horizontal comparative analysis of the level of higher education investment in the western region and the status quo of regional economic development (Li et al., 2020). Shen and Ho (2020) proposed a hybrid bibliometric method combining direct citation network analysis and text analysis to visually examine papers on higher education retrieved from the scientific network database (Shen and Ho, 2020). Sun (2020) analyzed the importance of psychological quality in college students' entrepreneurial education from the perspective of positive psychology and explored the educational mode of combining network information technology with maker education to cultivate college students' innovation and entrepreneurship ability (Sun, 2020). Godoy-Bejarano et al. (2020) provided evidence of how higher environmental complexity alters the slack-performance relationship in the long run by introducing three effects: efficiency effects, profitability effects, and incentive effects. After measuring these effects on the Colombian team, the results showed that the company had implemented a variety of simultaneous and purposeful actions against organizational slack to compete in a more complex environment (Godoy-Bejarano et al., 2020). Garland (2021) pointed out that positive psychology paid more attention to the positive qualities of human beings. The function of positive emotions is to expand and construct direct thoughts or behaviors of individuals, to provide sufficient resources for direct thoughts or behaviors of individuals, and to enable individuals to respond more accurately, perceive more comprehensively, and think more creatively (Garland, 2021). According to Guo et al. (2020), it is critical to implement positive psychology-based innovation and reform in college education. Network education resources play an important role in college education and are the most valuable form of teaching. However, the update speed of network resources is fast, and the utilization efficiency is low. Educational resources play an important role in schools, and educational resource recommendation plays a crucial mediating role in student learning (Guo et al., 2020). De Medio et al. (2020) proposed a fuzzy predictive control algorithm based on the Takagi and Sugeno (T–S) fuzzy model and integrated the nonlinear model into the neural network model to solve the problem of an insufficient linear relationship in the educational resource recommendation algorithm. It can be used to design educational resource recommendations (De Medio et al., 2020). Matos Pedro et al. (2020) indicated that intellectual capital in higher education institutions had a positive impact on institutional performance through relational capital and structural capital. Meanwhile, the quality of life has become an important aspect of performance standards in higher education institutions, especially in terms of students' perceptions of the quality of academic life. When higher education institutions understand and measure their intellectual capital, they will better understand their core competencies. This enables better allocation of resources and more effective strategic and operational actions (Matos Pedro et al., 2020). Li et al. (2022) focused on the use of big data and mobile computing-driven models to evaluate classroom teaching performance in their research. In addition, in the era of educational big data, their research also explored the general process of teachers' acquisition, analysis, and use of educational data to improve teaching performance. The data mining method and mobile data collection were organically combined into the benchmarking analysis. The classroom teaching performance of local colleges and universities was evaluated to enrich the teaching management theories and methods of local colleges and universities. The research results showed that benchmarking analysis could produce more meaningful results, which provided new data support for improving the quality of teaching management (Li et al., 2022).

Under the background of the transformation of China's economic development mode and the deepening of education and teaching reform, and under the background of the strategic development of a powerful country in education, this study combines the theory of positive emotions in positive psychology and the fuzzy predictive control algorithm to construct the T–S fuzzy neural network educational resource recommendation model. The application effect of the model is studied and analyzed. The innovation lies in the strategic level of strengthening the country through education, using the relevant theories of positive psychology, innovating in the way of recommending educational resources in resource colleges and universities, and constructing a teaching resource recommendation model for higher education management in the western region. The current situation of teaching and the feasibility and effect of model application are investigated and analyzed. The purpose is to improve the teaching management level of college students, improve the utilization rate and satisfaction of teaching resources, make teaching resources more widely used, and increase their utility.

2. Theoretical research and model method design

2.1. The strategic thought of education power and the theory of positive psychology

2.1.1. The influence of the strategic thought of powerful education on college education management

Building a strong country in education is the basic project for the great rejuvenation of the Chinese nation (Carr et al., 2021). General Secretary Xi Jinping has drawn a grand blueprint for the strategy of strengthening the country through education, planned a clear path, and provided scientific guidance and actions to follow. A strong country in education depends on whether the comprehensive strength, training ability, international competitiveness, and influence of education have a prominent position (Taufik, 2020). Adhering to education reform and innovation is the fundamental driving force for realizing the strategy of strengthening the country through education. If the construction of an educational power is not good, there will be insufficient intellectual resources to support international technological innovation and technological competitiveness. The essence of building a strong country in education is to improve the quality of education and promote equity in education (Li and Deng, 2022). The overall requirements of the new era for the strategic thinking of strengthening the country through education are reflected in four aspects, as shown in Figure 1.

First, people's satisfaction with education is the valuable goal of building an educational power. Second, giving priority to education is the basic requirement for building a strong educational country. Third, building morality and cultivating people is the fundamental task of building a powerful country through education (Renigier-Biłozor et al., 2019). Fourth, Chinese characteristics and world-class education are the scientific orientation for developing a powerful country. Based on the strategic thinking of strengthening the country through education, and under the impetus of the new round of the western development policy, the state has provided "blood-transfusion" support for higher education in the western region, continuously supplemented the shortcomings in human, financial, and material resources, and continuously built an industrial service chain connecting the inside and outside of the western region to enhance the conversion rate of scientific research achievements in colleges and universities (Dana et al., 2021).

As an important support for the construction of a strong country, higher education is a key area for the development of the western region and an overall boost to the economy and society in the western region. Integrating educational resources on the basis of the industrial service chain can not only improve the effect of collaborative education, but also provide development opportunities for forming the characteristics of western universities and adjusting the local industrial structure. Under the strategic thinking of strengthening the country through education, the western region should coordinate and integrate innovation, system innovation, and institutional innovation to realize fundamental changes in the way resources are acquired and the transformation of scientific research results.

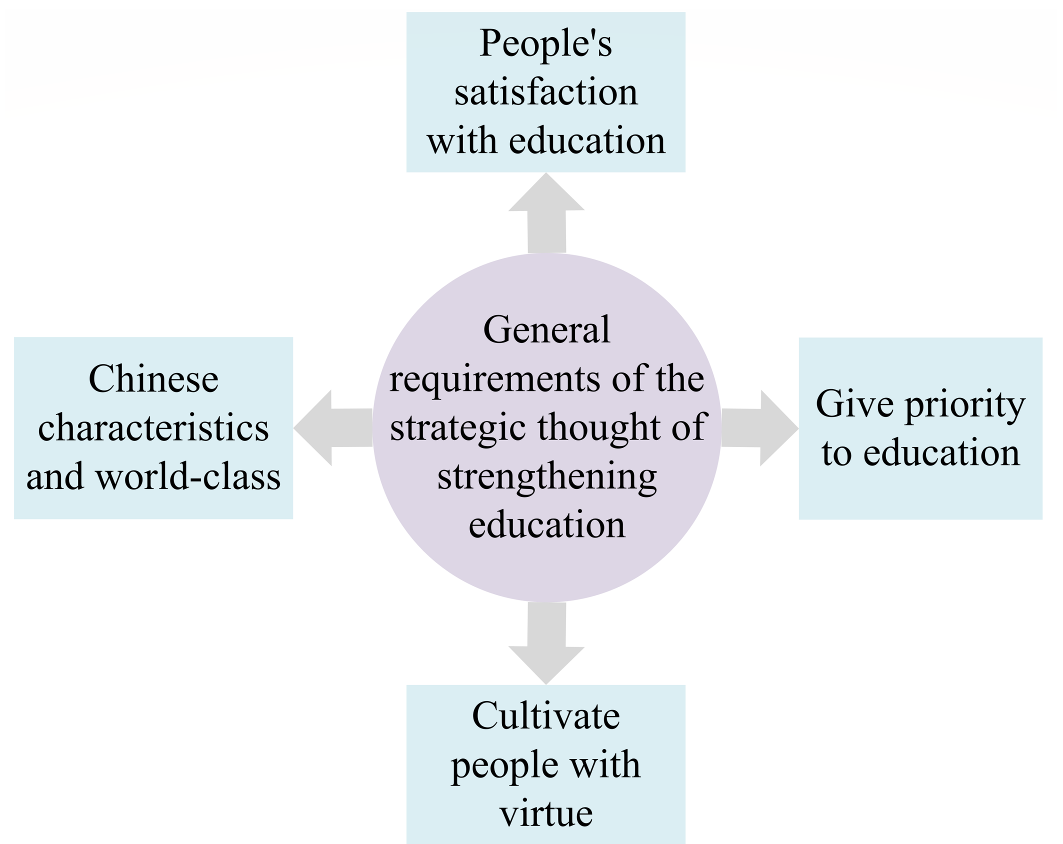


FIGURE 1
Reflection of the general requirements of the strategic thinking of strengthening the country through education.

2.1.2. The effect of positive psychology theory on educational management In colleges and universities

Positive psychology is to use the existing, relatively complete psychological research methods to explore the development of human potential, the satisfaction of needs and the improvement of psychological quality to finally obtain a happy life. Positive psychology emphasizes putting people first and advocates human care for all students. Its purpose is to stimulate and cultivate students' individual constructive strength and positive psychological quality, guide students to give full play to their self-help and self-healing ability when they encounter problems and difficulties, and finally realize students' self-development (Baig et al., 2021). Secondly, positive psychology attaches great importance to prevention and promotion of growth, and advocates that, in the face of complex student work problems, individuals should be guided to use their own positive qualities and strengths to stop the problems in their infancy. Positive psychology enables students to understand their own strengths and weaknesses more comprehensively and accurately, and rebuild self-confidence (Kim et al., 2021).

Positive emotional experience is a branch of research in positive psychology. The function of positive emotions is the ability to expand and structure a person's direct thoughts or behaviors. Positive emotions provide sufficient resources for an individual's direct thoughts or actions, enabling individuals to respond more accurately, to be more comprehensive, and to think creatively. In addition to schools and families, the cultivation of college students' self-management ability is more important than the role of their own psychology (Wang et al., 2022). Therefore, positive psychology plays an important role in higher education management and student teaching. Positive emotional experiences can not only reflect personal happiness, but also benefit personal growth and development. Positive emotions can expand students' thinking and action (Rodríguez-Campo et al., 2022). The specific extension construction is shown in Figure 2.

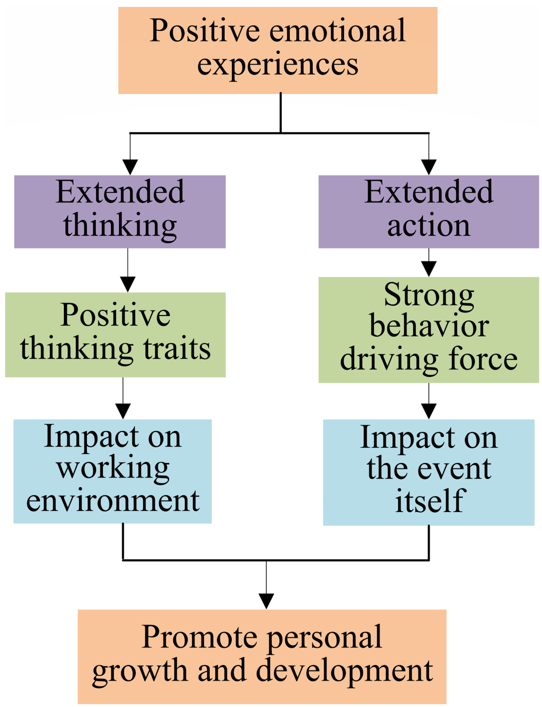


FIGURE 2
An extended theory of positive emotions.

In Figure 2, positive emotional experiences provide an opportunity to build sustainable personal resources. Positive emotional experience shows people's positive thinking characteristics and strong behavioral drive. This kind of emotion affects people's work attitudes and events and then influences the work environment and the results of work events. Positive emotions provide the potential for personal growth and development.

2.2. Construction of higher education management resource recommendation model based on T–S fuzzy neural network

2.2.1. Relationship between recommendation mode of educational management resources and positive psychological emotion and design of fuzzy control algorithm

2.2.1.1. The relationship between the recommended mode of educational resource management and positive psychological emotion

The first priority in the advancement of teaching work in colleges and universities is the optimization of teaching power strategy, the most important of which is the acquisition and application of educational management resources. The recommendation system for educational management resources in colleges and universities includes modules such as acquisition of educational resources, educational resources, recommendation resources, personal educational resources, and subject information management. Among them, the manner in which educational resources are obtained has changed dramatically in recent years. From the traditional book acquisition method to the current information acquisition method, the great change in the method of obtaining educational resources has also had a great impact on educational enthusiasm. Educational resources are also called "educational economic conditions." The educational process occupies, uses, and consumes human, material, and financial resources, that is, the sum of educational human resources, material resources, and financial resources. Human resources include educators' and educators' human resources, that is, the number of students in a school, class, enrollment, graduates, administrators, teaching staff, teaching assistants, workers, production staff, etc. Material resources include fixed assets, materials, and low-value consumables in schools. Fixed assets are divided into common fixed assets, fixed assets for teaching and scientific research, and other general equipment fixed assets. Recommending resources refers to the formation of an educational resource recommendation system through information technology, so as to improve the acquisition effect of teaching resources. Specialty and skill resources, professional resources, personal contacts resources, regional resources, and so on are all examples of personal educational resources (Xiong et al., 2021). Discipline management entails implementing discipline management using information technology and improving the effectiveness of discipline management.

To sum up, the current recommended mode of educational resource management is basically realized by information technology, so it can improve the effect of educational resource management, thereby helping teachers and learners to improve their enthusiasm. Meanwhile, constantly updated educational resource management methods and contents can effectively stimulate educators' and learners' enthusiasm in teaching and learning, thereby comprehensively promoting the development of teaching work. Based on this, this paper combines the recommended mode of educational resources management with positive psychological and emotional factors to study the teaching work in colleges and universities to promote the development of teaching work. Figure 3 shows the basic idea of this research.

In Figure 3, under the background of the optimization of teaching power strategy, the optimization of teaching resource management methods and the promotion of enthusiasm can effectively promote the development of teaching work, thus enhancing the development effect of current colleges and universities.

2.2.1.2. Fuzzy control algorithm design

The fuzzy control algorithm is essentially a kind of computer numerical control. It consists of fuzzy set theory, fuzzy linguistic variables, and fuzzy logic reasoning (Kong et al., 2019). The fuzzy predictive control algorithm combines the advantages of the fuzzy and the predictive control algorithm and combines the fuzzy ideas and prediction methods of the two. The effect of predictive control is enhanced through fuzzy prediction, feedback correction, and rolling optimization (Thamallah et al., 2019). Among the forms of fuzzy sets, there are three common ones. The first is the Zadeh form (Lin et al., 2021). The fuzzy set A is shown in Eq. 1:

$$A = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \cdots + \frac{\mu_A(x_n)}{x_n} \tag{1}$$

$\frac{\mu_A(x_1)}{x_1}$ represents the correspondence between the element x_1 and the degree of membership $\mu_A(x_1)$. In ordinal even form, the expression of fuzzy set A is shown in Eq. 2:

$$A = \{(x_1, \mu_A(x_1)), (x_2, \mu_A(x_2)), \cdots, (x_n, \mu_A(x_n)) \mid x \in U\} \tag{2}$$

In vector form, the fuzzy set A is given by Eq. 3:

$$A = [\mu_A(x_1)\mu_A(x_2)\cdots\mu_A(x_n)] \tag{3}$$

In Eqs 1–3, when the degree of membership is 0, this term is omitted in the Zadeh form and the ordinal pair form. In vector form, this item cannot be omitted. In practice, membership function is a key part of fuzzy set theory, which quantitatively describes fuzzy concepts. The general fuzzy distribution functions are divided into normality, upper and lower ring types (Anil Naik, 2022).

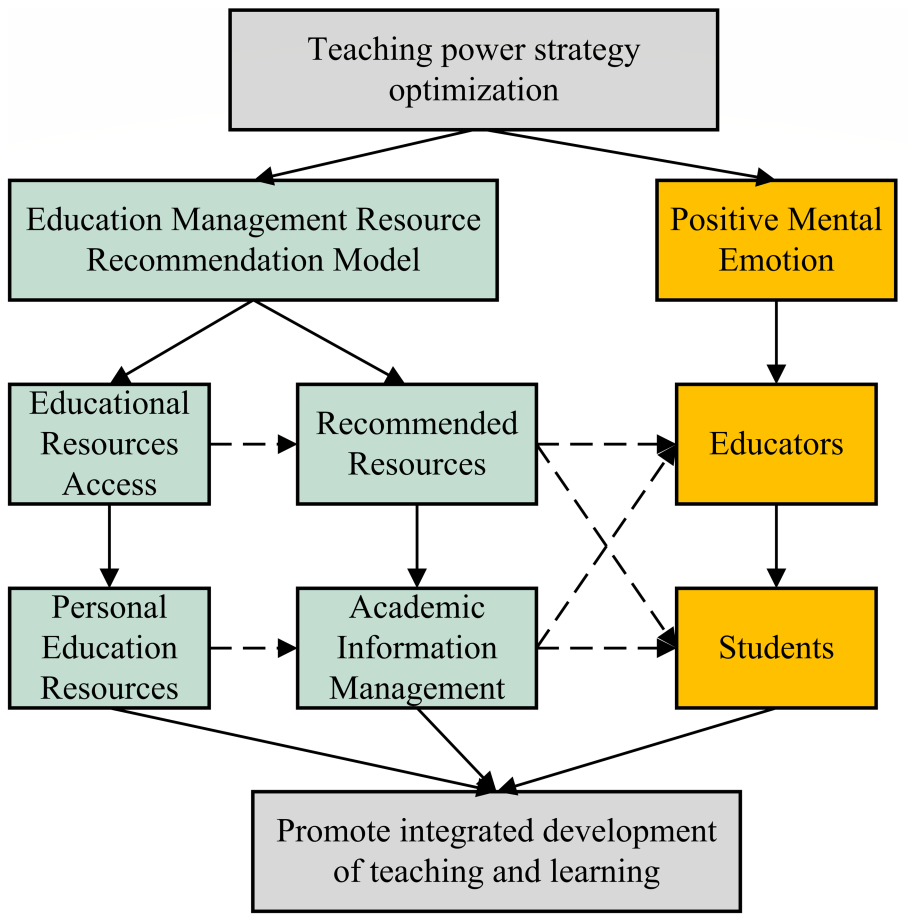


FIGURE 3
Design of research ideas.

2.2.2. T–S fuzzy model rule design

Takagi and Sugeno (T–S) was created by Takagi and Sugeno, referred to as the T–S fuzzy model. The essence is to divide the input space into multiple fuzzy subspaces and build a linear relationship model between input and output in the subspace, following the If-then rule (Zeng et al., 2019).

In the fuzzy model, the input vector x is shown in Eq. 4:

$$x = \left[x_1, x_2, \cdots x_i, x_n\right]^T \tag{4}$$

Among them, x_i is a fuzzy language variable. The expression of $T(x_i)$ is shown in Eq. 5:

$$T(x_i) = \left\{A_i^1, A_i^2, \cdots, A_i^{m_i}\right\}(i = 1,2,\cdots,n) \tag{5}$$

$A_i^j\,(j = 1,2,\cdots,m_i)$ represents the *j*-th language variable value of x_i, representing a fuzzy set of U_i. The expression of the membership function is shown in Eq. 6:

$$\mu_{A_i^j}(x_i)(i = 1,2,\cdots,m_i) \tag{6}$$

T–S Fuzzy Model Rule Consequences represent linear combinations of input variables. Suppose x_1 is A_1^j, x_2 is A_2^j. Then, x_n is A_n^j. The expression of y_j is shown in Eq. 7:

$$y_j = p_{j0} + p_{j1}x_1 + \cdots p_{jn}x_n \tag{7}$$

Among them, $j = 1,2,\cdots,m$, $m \le \prod_{i=1}^{n} m_i$.

If the input quantity uses the fuzzification method of a single-point fuzzy set, for the given input x, the rule fitness α_j is obtained, as shown in Eq. 8:

$$\alpha_j = \mu_{A_1^j}(x_1) \wedge \mu_{A_2^j}(x_2) \cdots \wedge \mu_{A_n^j}(x_n) \tag{8}$$

The output of the fuzzy system is the weighted average of the output of each rule, represented by y, as shown in Eq. 9:

$$y = \frac{\sum_{j=1}^{m}\alpha_j y_j}{\sum_{j=1}^{m}\alpha_j} = \sum_{j=1}^{m}\bar{\alpha}_j y_j \tag{9}$$

$\bar{\alpha}_j$ is shown in Eq. 10:

$$\bar{\alpha}_j = \alpha_j \Big/ \sum_{i=1}^{m}\alpha_i \tag{10}$$

Suppose: there are m rules. Then, the *i*-th rule is shown in Eq. 11:

$$R_i : \text{If } x(t) \text{ is } A_i \text{ then } y_i(t+1) = a_{i0} + a_{i1}x_{t1} + \cdots + a_{im}x_{tm} i = 1,2,\cdots,m \tag{11}$$

In fuzzy rules, $x(t)$ is shown in Eq. 12:

$$x(t) = \left[x_{t1}, x_{t2}, \cdots x_{ts}\right] \tag{12}$$

In the If-then rule, the part before the then statement is called the antecedent; A_i is a fuzzy set composed of regression vectors; the part after the then statement is called If. Then rule consequences represent linear combinations of input and output data. In the If-then rule, the data information of the rule's consequent can only be obtained if the antecedents of the then rule are satisfied (Yi et al., 2021).

The least-squares method is used to resolve the regular consequent to obtain a_{im}. If the membership degree of $x(t)$ to the *i*-th rule is $\mu_i(t)$, the output is shown in Eq. 13:

$$\hat{y}(t) = \sum_{i=1}^{m}\beta_i(t)y_i(t) \tag{13}$$

$\beta_i(t)$ is shown in Eq. 14:

$$\beta_i(t) = \mu_i(t) \Big/ \sum_{i=1}^{m}\mu_i(t) \tag{14}$$

2.2.3. Design of recommendation process for higher education management resources based on T–S fuzzy neural network

Previously, educational resource management research focused on the analysis and discussion of the current state of educational resources, as well as recommendations for improving the current state of educational resources (Li and Sun, 2018). In addition, the research on the optimization of higher education management and the improvement of management strategies by technical methods is more detailed, and the research on the construction of an information platform in the recommendation for college educational resources is relatively lacking. The system of college education resource recommendation includes modules such as educational resource acquisition, educational resources, recommended resources, personal education resources, and subject information management. A T–S fuzzy neural network structure based on educational resource recommendation is constructed (Lu et al., 2020).

Feedforward neural network has a multilayer perceptron structure. The multilayer perceptron treats each input data as a vector $x = (x_1, x_2)$, and obtains the transfer function (Ozanich et al., 2020), as shown in Eq. 15:

$$f(x) = x \cdot w + b \tag{15}$$

w is the weight vector, and b is the vertical offset (Gallab et al., 2019).

Nonlinear factors are introduced. The perceptron can approximate any nonlinear function, and the activation function activates the transfer function. Finally, the result is shown in Eq. 16:

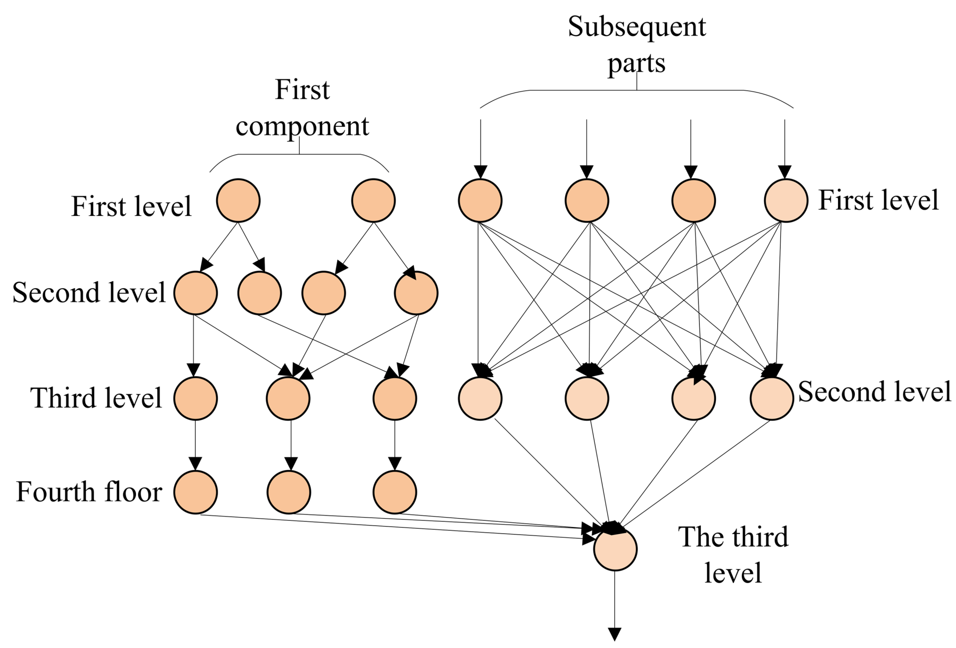


FIGURE 4
The structure of a multilayer perceptron.

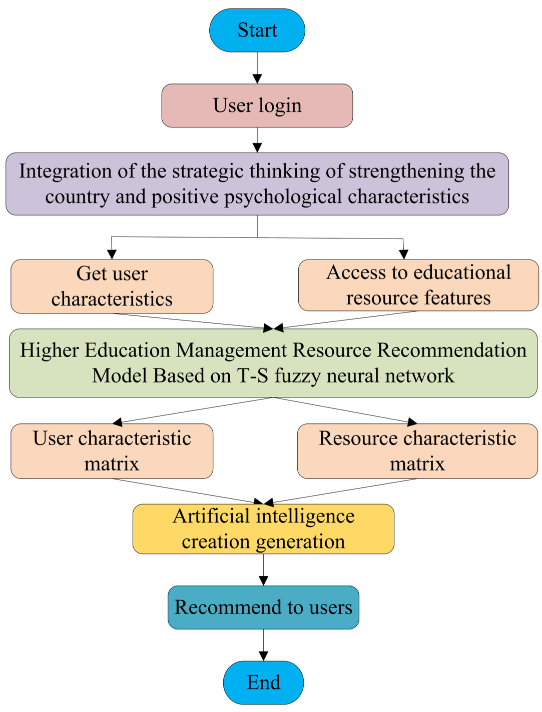


FIGURE 5
T–S fuzzy neural network educational resource recommendation model process.

$$h(x) = \begin{Bmatrix} 1: \text{if } f(x) = x \cdot w + b > 0 \\ 0: \text{others} \end{Bmatrix} \tag{16}$$

Based on the theory of fuzzy generalized predictive control based on T–S model, a higher education management resource recommendation model based on T–S fuzzy neural network is constructed. The details are shown in Figure 4.

In Figure 4, the neural network consists of an input, an output, and multiple hidden layers. Every neuron is a perceptron. The neurons of the input layer serve as the input to the hidden layer. The hidden layer's neurons are used as the input of the output layer, and the vector is input to the input layer. After repeated operations in the hidden layer, the output layer finally outputs the result of the neural network.

2.3. Recommendation process of higher education management resources based on T–S fuzzy neural network

This design is based on a neural network and fuzzy control algorithm, combined with the strategic thinking of strengthening the country through education, and under the action of positive psychology, a higher education management resource recommendation model based on a T–S fuzzy neural network is constructed, and the model process design is further carried out (Yao and Shekhar, 2021). The details are shown in Figure 5.

Figure 5 shows that starting from the user login, based on the strategic thinking of strengthening the country through education and the characteristics of positive psychology, through the data of user characteristics and educational resource characteristics, a T–S fuzzy neural network-based higher education management resource recommendation model is constructed. Further, a user feature matrix and an educational resource feature matrix are formed. The model recommends the most suitable resources to users through repeated calculations and predictions, and after multiple feedback corrections, and the entire recommendation process ends.

2.4. Investigation method and model demonstration

The first part is a simulation experiment of the educational resource recommendation design of the T–S fuzzy neural network. The second part is a questionnaire survey on the current situation of educational resources in colleges and universities and the application of educational resource recommendation design. The results before and after the design application are compared and analyzed.

2.4.1. Experimental test

The system is tested by laboratory simulation, which verifies the feasibility of the fuzzy predictive control model based on neural networks in the recommendation of educational resources. Test requirements: the recommended educational resources must conform to the characteristics of users. That is, they must be related to the actual needs of users, and the recommendation accuracy rate should be >85%. In the experiment, the model is first trained and evaluated, and then tested and evaluated. Among them, the training evaluation time is 2 h, the training times of the model are 1,000 times, and the convergence time of the model is 1 h.

The test environment is shown in Table 1.

2.4.2. Design part of the questionnaire

Through the method of a questionnaire, taking the students of M colleges and universities in Xinjiang in the western region as research objects, the basic situation of students' learning and teaching, and the effect of teaching resource recommendation design and application are investigated. The current situation of teaching resources, the influence of positive psychology on teaching effects, and the teaching situation before and after the application of the educational resource recommendation model for higher education management are investigated. In order to ensure the authenticity and reliability of the questionnaire content, the validation of the questionnaire is tested. Six education experts are invited to give a comprehensive evaluation of the rationality of the questionnaire content design. The specific results are shown in Table 2.

The second part is the distribution and collection of questionnaires. At M universities, a questionnaire survey was administered to freshmen, sophomores, juniors, and some teachers. This survey adopts a small program of asking questions and using paper questionnaires. A total of 300 questionnaires were distributed, 290 were collected, and 280 were valid, for an effective rate of 96.6%. Statistical Product Service Solutions (SPSS) 22.0 was used to sort out, count, and analyze the questionnaires. Among them, 66.67% of the questionnaires are very reasonable, and 33.33% are relatively reasonable. In this paper, the reliability and validity of the questionnaire are tested, and the reliability coefficient of the questionnaire is 0.860 and the validity coefficient is 0.710. The results show that the questionnaire design is reasonable.

TABLE 1 Settings for the test environment.

Configure	Information
Laptop	Dell Inspiron 3,420
Operating system	Window10 64
Processor	Intel(R) Core (TM) i5-321OM
Memory	8G
Database	MySQL Server 5.6

3. Test results and investigation analysis

3.1. Analysis of the current situation of the investigation of educational management resources in colleges and universities

3.1.1. Analysis of the current situation of teaching faculty

In order to understand the details of the teaching staff in detail, the three aspects of the teaching staff's professional title, education background, and age are analyzed. All personnel are divided into full-time and part-time, as shown in Figure 6.

In Figure 6, full-time teachers account for the largest proportion of lecturers, accounting for 45.65%, followed by associate professors, accounting for 44.66%, professors, 16.1%, and teaching assistants, 10.52%. Among the part-time teachers, associate professors account for the largest proportion, accounting for 44.66%, followed by professors, accounting for 28.42%. The overall proportion of school professors is relatively low. In terms of academic qualifications, master's degree is the main force among full-time teachers, accounting for 58.62%, followed by doctors, accounting for 27.49%, and some undergraduates, accounting for 13.89%. Among the part-time teachers, doctors are the most common, accounting for 79.76%. It shows that the academic qualifications of full-time teachers in the school are generally not high. From the perspective of age composition, most of them are under 40 years old, followed by those over 50 years old. There are not many full-time, young and experienced teachers. Full-time teachers are not strong enough.

3.1.2. Analysis of the basic situation of school professional resources

The professional resources of the school are analyzed based on the proportion of key majors, characteristic majors, and general majors among the different disciplines of economics, management, engineering, science, law, and literature. The results are shown in Figure 7.

In Figure 7, in the situation of professional resources, the number of key majors is relatively small, the proportion of key majors and characteristic majors in economics and management disciplines is slightly higher, and other disciplines are generally lower. The school has a relatively large proportion of general majors. Professional advantages do not have great competitiveness, and the advantages are not obvious.

3.2. Feasibility analysis of application of higher education management resource recommendation model

According to the educational resource recommendation model, for the educational resources recommended by the system platform, including reference books, research articles, online teaching resources, and the latest educational information, the accuracy of educational resource recommendation is compared before and after, and the results are shown in Figure 8.

Figure 8 shows that after the application of this model, the recommendation accuracy of reference books, research papers, online

TABLE 2 Statistical table for questionnaire validity test.

Validity	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	frequency	percent (%)
Very reasonable	√		√	√		√	4	66.67
More reasonable		√			√		2	33.33
Not very reasonable							0	0
Unreasonable							0	0

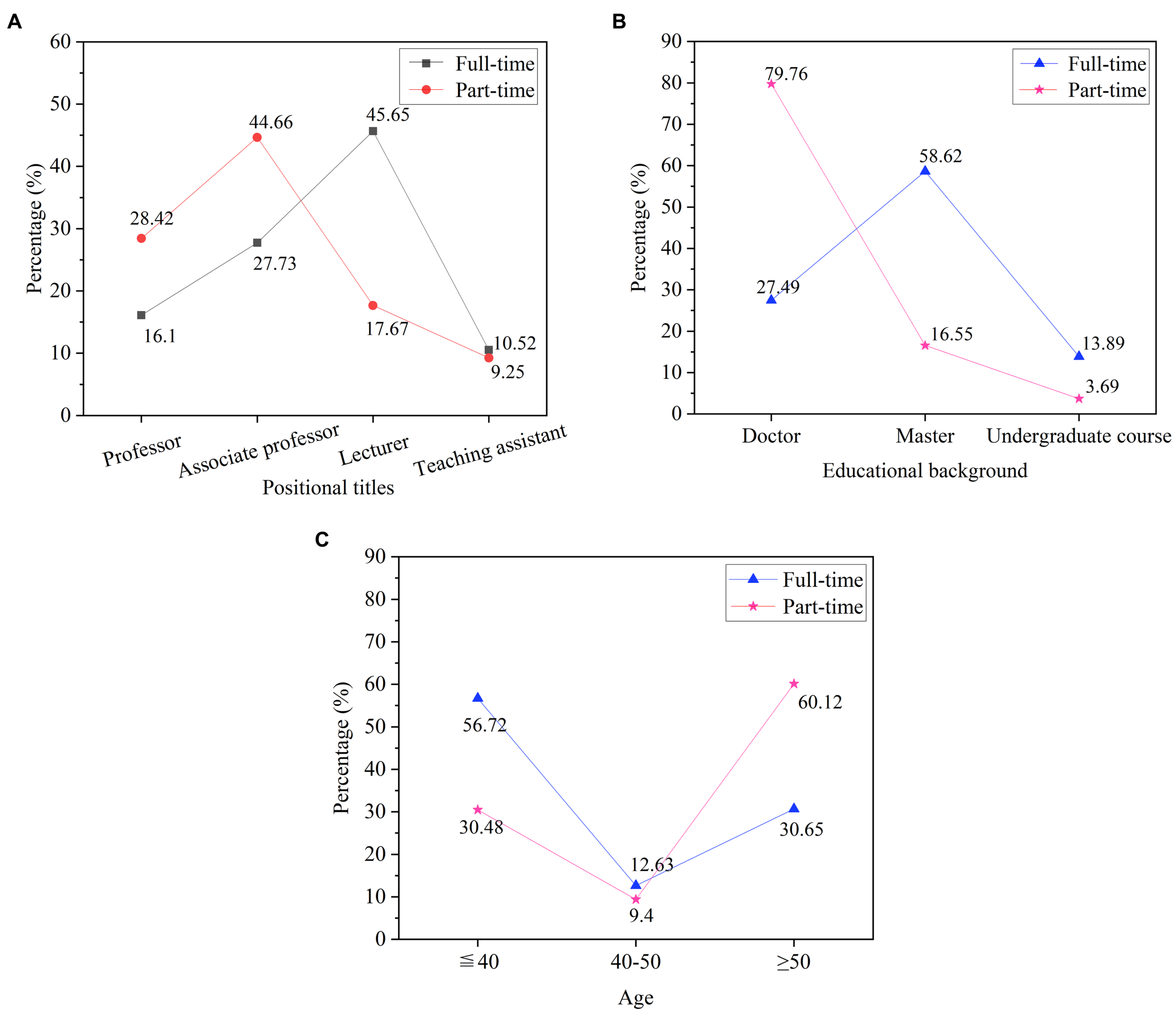


FIGURE 6
Current status of teaching faculty. **(A)** the professional titles; **(B)** the academic qualifications; **(C)** the age.

teaching resources, and the latest educational information recommended by the platform has improved. The recommendation accuracy of reference books increased from 80.2% to 84.3%, that of research articles from 83.3% to 85.6%, that of online teaching resources from 85.9% to 95.1%, and that of the latest educational information from 79.4% to 86.8%. The accuracy rate has been improved on the whole. This shows that the design model is feasible. In addition, this paper also evaluates the effect of the model in the process of recommending actual educational resources, as shown in Figure 9, which shows the results of this paper's evaluation of the effect of the model in recommending history and mathematics curriculum resources.

In Figure 9, the model designed in this paper performs very well in the research of curriculum resource recommendations in colleges and universities. The highest fitting degree of the model in the recommendation of teaching management resources for history and mathematics courses is about 83%, and the lowest is about 75%. It can be seen that the model designed in this paper has a very good effect on the recommendation of teaching management resources.

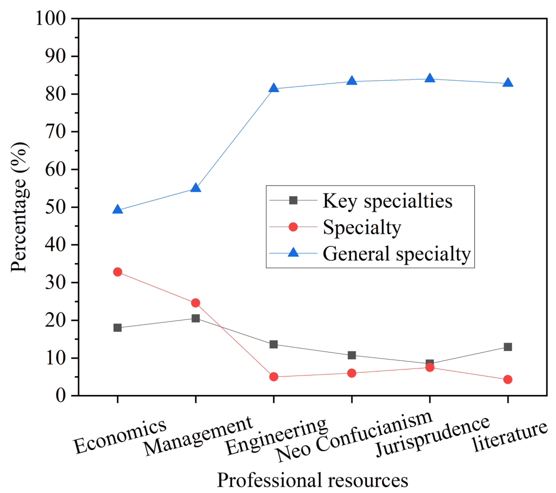


FIGURE 7
Analysis of professional resources in different disciplines.

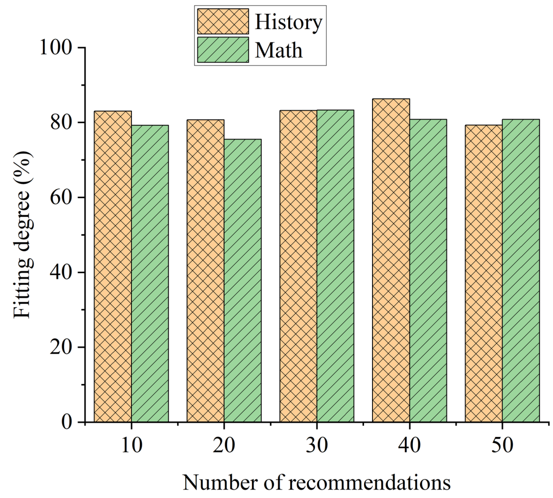


FIGURE 9
Evaluation of recommendation effect of model educational resources.

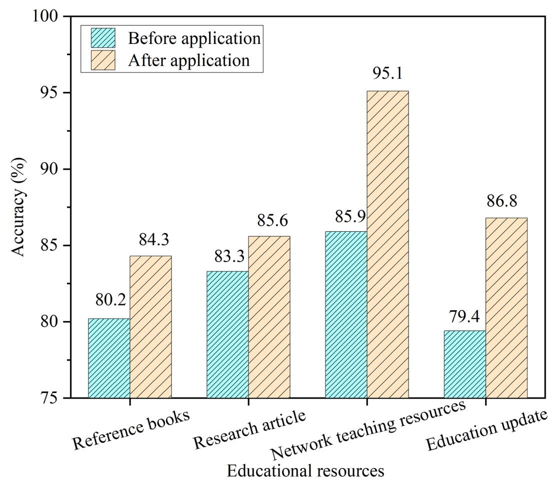


FIGURE 8
Comparison of educational resource recommendation results before and after model application.

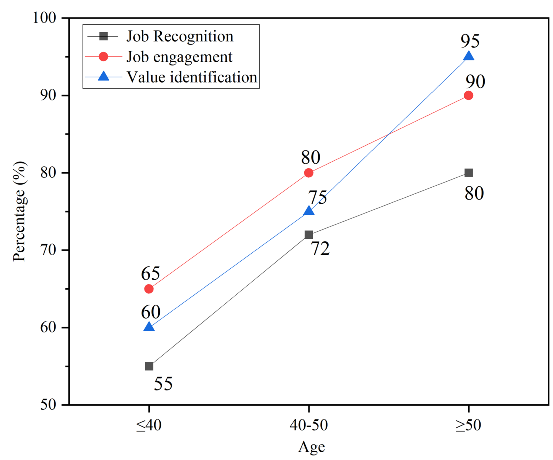


FIGURE 10
The influence of psychological emotions on teachers' work attitude.

3.3. Analysis on the effect of higher education teaching management integrating positive psychological emotions

3.3.1. The influence of positive psychological emotions on teachers' work attitude

For teachers of different ages, the influence of positive psychological emotions on teachers' work recognition, work engagement and value recognition in education and teaching management are analyzed, as shown in Figure 10.

In Figure 10, older teachers have higher recognition of their own work and are more engaged in their work. The stronger the sense of value recognition of the work, the deeper the qualifications and the longer the working hours, the more contributions to the society, the more positive emotions affect them, and the more positive work attitudes are promoted. Young people need to have a more positive attitude toward work to meet new tasks and improve their awareness of the value of work.

3.3.2. The influence of positive psychological emotions on teachers' psychological state

For teachers of different ages, the impact of positive psychological emotions on personal vitality, dedication, and concentration is analyzed, as shown in Figure 11.

In Figure 11, among teachers under the age of 40, the personal vitality is still relatively strong, and the vitality gradually decreases with age. In terms of dedication, teachers over the age of 50 have a stronger sense of commitment, followed by teachers aged 40–50, and finally, teachers under the age of 40. The gap in concentration is not very big, they can maintain good concentration, and teachers over 50 are slightly better. Positive psychological emotions can promote and enhance teachers' dedication and concentration to a certain extent.

3.3.3. The influence of positive psychological emotions on students' conflict handling

The interpersonal aspect is mainly manifested in the communication and conflict resolution among the classmates. Rational resolution and calm analysis of problems and contradictions among college students

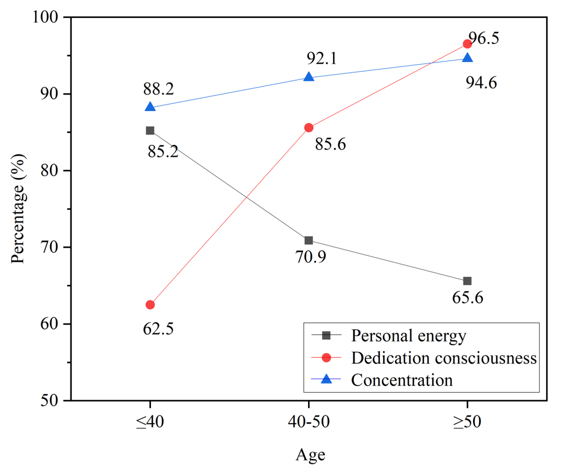


FIGURE 11
The influence of psychological emotions on teachers' psychological state.

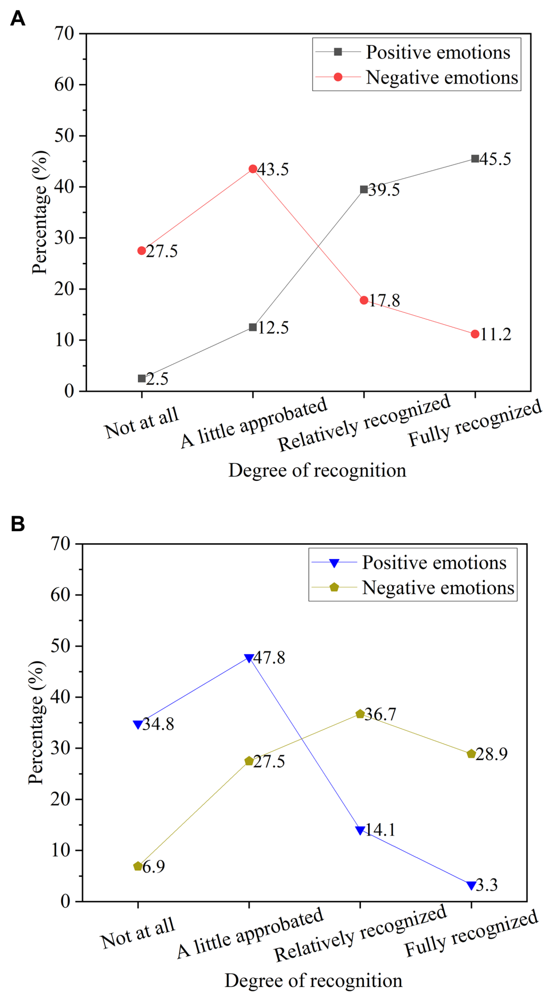


FIGURE 12
The influence of psychological emotions on the problem solving of college students. **(A)** The recognition of different psychological emotions for calmly analyzing and resolving conflicts; **(B)** The recognition of different psychological emotions for intensifying conflicts and behavioral confrontation.

under psychological and emotional conditions and recognition of behaviors and confrontations caused by intensified conflicts are analyzed, as shown in Figure 12.

In Figure 12, after conflicts arise among students, more students will calmly analyze and rationally resolve conflicts with positive psychological emotions. Under a negative psychological mood, the proportion of students' ability to calmly analyze and rationally resolve conflicts decreases. A positive psychological state has a positive effect on conflict resolution. Positive psychological emotions can reduce the possibility of intensifying conflict and the possibility of behavioral confrontation.

3.4. Analysis of the impact of the application of the teaching resource recommendation model on the teaching effect

3.4.1. Satisfaction analysis of query results of teaching resources under model application

Comparing the acceptance of the teaching resource query results and the satisfaction of the query efficiency of college students before and after the application of the teaching resource recommendation model, the results are shown in Figure 13.

Figure 13 shows that after the design and application of the recommended reference books, research articles, online teaching resources and education latest information, the acceptance of the query results and the satisfaction of query efficiency of the four types of resource recommendation have improved, especially for the query results of online teaching resources. The acceptance rate is the strongest, with the proportion increasing from 57.6% to 90.2%. In terms of query efficiency, satisfaction with the query efficiency of online teaching resources is the highest, ranging from 59.8% to 85.6%. Overall, college students are satisfied with the application of the teaching resource recommendation model in resource query.

3.4.2. Satisfaction analysis of teaching resource application under model application

The degree of satisfaction of college students with the application of teaching resources before and after the recommended design and application of teaching resources is compared and analyzed, and the results are shown in Figure 14.

Figure 14 shows that after the application of the teaching resource recommendation model, the proportion of resource search convenience has increased, from 53.9% to 70.5%. The percentage of people who thought it was not helpful after the application dropped, from 10.8% to 4.2%. The percentage of people who felt the system had limitations fell from 17.1% to 9.6%. The teaching resource recommendation model can enhance students' interest in the application of teaching resources as a whole, and application satisfaction is significantly improved.

4. Conclusion

In order to improve the level of higher education management in the western region, this study first conducts an in-depth study on the strategic thought of strengthening the country through education and the impact of positive psychology on higher education management, and then uses the T–S fuzzy neural network model to construct an educational resource recommendation based on the T–S fuzzy neural network. The model is designed, the process design is carried out, and the model is verified and

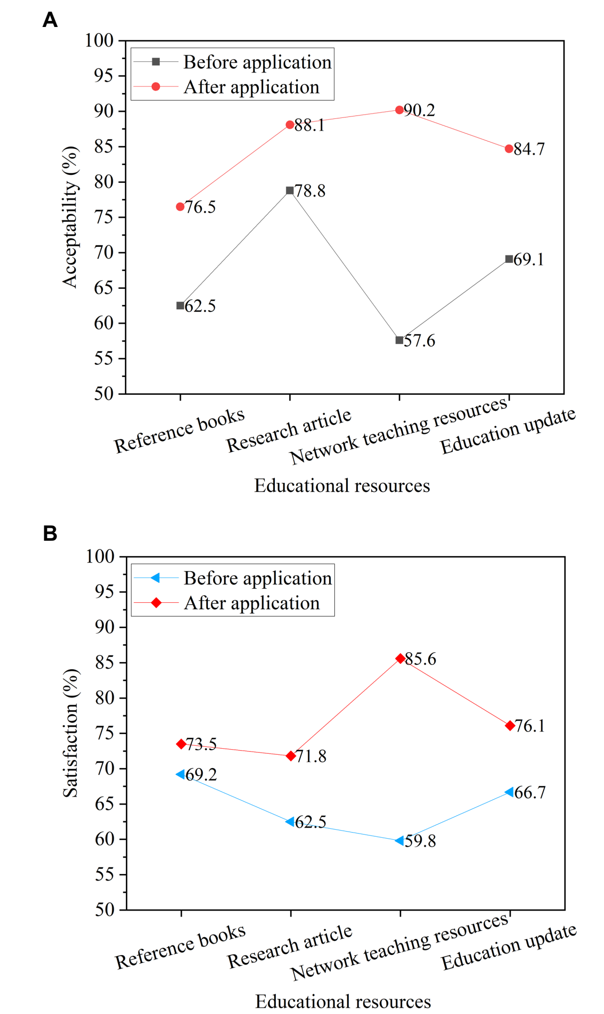


FIGURE 13
Satisfaction of teaching resource query results. **(A)** Comparison of acceptance of teaching resource query results; **(B)** Satisfaction comparison of teaching resource query efficiency.

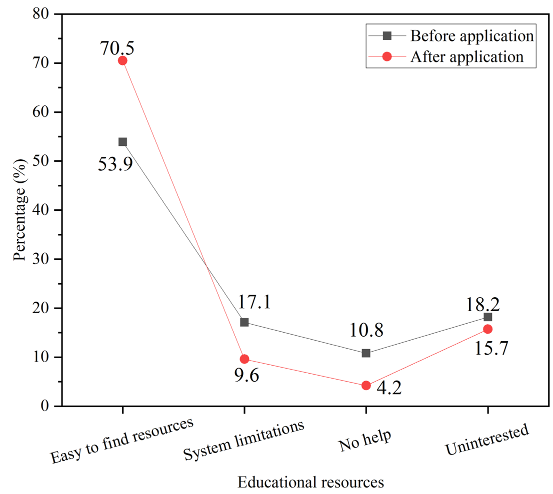


FIGURE 14
Comparison of teaching resource application satisfaction.

tested. A questionnaire survey is used to conduct a comprehensive analysis of the current state of educational resources in colleges and universities, as well as the application effect of the resource recommendation model. The results show that: (1) When analyzing the current situation of the M College Education Resources Survey, the overall education of full-time teachers is not high, and the proportion of young full-time teachers with certain experience is relatively small. The school has a large proportion of ordinary majors, and the professional advantages are not competitive and not obvious. (2) After the model is applied, the recommendation accuracy of the four types of educational resources has improved, and the design is feasible. (3) Positive psychological emotions have a significant impact on the effectiveness of higher education management teaching. Positive psychological emotions can greatly enhance teachers' sense of dedication and concentration, and positive psychological emotions can reduce the possibility of intensifying conflicts and behavioral confrontation. (4) College students are generally satisfied with the teaching resource recommendation model. The teaching resource recommendation model can enhance college students' interest in the application of teaching resources, and application satisfaction is significantly improved. The limitation of this research is that, due to limited ability, the scope of the survey object is not broad enough, and future research should broaden the scope of the research and conduct experimental and popularized research in other universities in the western region. The main contribution is to use the relevant theories of positive psychology to construct the recommendation of higher education management resources based on T–S fuzzy neural networks in the western region, to investigate and analyze the teaching status and the feasibility and effect of the model application, and to improve the research on the recommendation system of educational resources in educational management, which improves the recommendation effect of educational resources.

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

XMS contributed to conception and design of the study, organized the database and performed the statistical analysis, wrote the first draft of the manuscript, and wrote sections of the manuscript. XMS contributed to the manuscript revision, read, and approved the submitted version.

Funding

This work was supported by the High-level Talents Scientific Research Initiation Project of Shihezi University (no. RCSK202001), Shihezi University Youth Innovation and Cultivation Talent Project (no. CXPYSK202103), and Independent Financial Support Project of Shihezi University (no. ZZZC202168).

Conflict of interest

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

Publisher's note

References

Anil Naik, K. (2022). Real time implementation of interval type-2 fuzzy logic based damping controller for wind integrated power system. *Wind Eng.* 46, 392–412. doi: 10.1177/0309524X211030186

Baig, S. A., Iqbal, S., Abrar, M., Baig, I. A., Amjad, F., Zia-ur-Rehman, M., et al. (2021). Impact of leadership styles on employees' performance with moderating role of positive psychological capital. *Total Qual. Manag. Bus. Excell.* 32, 1085–1105. doi: 10.1080/14783363.2019.1665011

Carlucci, D., Renna, P., Izzo, C., and Schiuma, G. (2018). Assessing teaching performance in higher education: a framework for continuous improvement. *Manag. Decis.* 57, 461–479. doi: 10.1108/MD-04-2018-0488

Carr, A., Cullen, K., Keeney, C., Canning, C., Mooney, O., Chinseallaigh, E., et al. (2021). Effectiveness of positive psychology interventions: a systematic review and meta-analysis. *J. Posit. Psychol.* 16, 749–769. doi: 10.1080/17439760.2020.1818807

Dana, L. P., Tajpour, M., Salamzadeh, A., Hosseini, E., and Zolfaghari, M. (2021). The impact of entrepreneurial education on technology-based enterprises development: the mediating role of motivation. *Administrative Sci.* 11:105. doi: 10.3390/admsci11040105

De Medio, C., Limongelli, C., Sciarrone, F., and Temperini, M. (2020). Moodle REC: a recommendation system for creating courses using the moodle e-learning platform. *Comput. Hum. Behav.* 104:106168. doi: 10.1016/j.chb.2019.106168

Dong, Q. W., Wang, S. M., Han, F. J., and Zhang, R. D. (2019). Innovative research and practice of teachers' teaching quality evaluation under the guidance of 'innovation and entrepreneurship'. *Procedia Comput. Sci.* 154, 770–776. doi: 10.1016/j.procs.2019.06.123

Dopson, S., Ferlie, E., McGivern, G., Fischer, M. D., Mitra, M., Ledger, J., et al. (2019). Leadership development in higher education: a literature review and implications for programme redesign. *High. Educ. Q.* 73, 218–234. doi: 10.1111/hequ.12194

Fu, W., Tang, W., Xue, E., Li, J., and Shan, C. (2021). The mediation effect of self-esteem on job-burnout and self-efficacy of special education teachers in Western China. *Int. J. Dev. Disabilities* 67, 273–282. doi: 10.1080/20473869.2019.1662204

Gallab, M., Bouloiz, H., Alaoui, Y. L., and Tkiouat, M. (2019). Risk assessment of maintenance activities using fuzzy logic. *Procedia Comput. Sci.* 148, 226–235. doi: 10.1016/j.procs.2019.01.065

Garland, E. L. (2021). Mindful positive emotion regulation as a treatment for addiction: from hedonic pleasure to self-transcendent meaning. *Curr. Opin. Behav. Sci.* 39, 168–177. doi: 10.1016/j.cobeha.2021.03.019

Godoy-Bejarano, J. M., Ruiz-Pava, G. A., and Téllez-Falla, D. F. (2020). Environmental complexity, slack, and firm performance. *J. Econ. Bus.* 112:105933. doi: 10.1016/j.jeconbus.2020.105933

Guo, J., Li, C., Zhang, G., Sun, Y., and Bie, R. (2020). Blockchain-enabled digital rights management for multimedia resources of online education. *Multimed. Tools Appl.* 79, 9735–9755. doi: 10.1007/s11042-019-08059-1

Kim, E. J., Kim, J. J., and Han, S. H. (2021). Understanding student acceptance of online learning systems in higher education: application of social psychology theories with consideration of user innovativeness. *Sustainability* 13:896. doi: 10.3390/su13020896

Kong, L., He, W., Yang, C., Li, Z., and Sun, C. (2019). Adaptive fuzzy control for coordinated multiple robots with constraint using impedance learning. *IEEE Trans. Cybern.* 49, 3052–3063. doi: 10.1109/TCYB.2018.2838573

Li, Y., and Deng, M. (2022). Research on the innovation and development of university physical education from the perspective of strengthening the country through sports. *Creat. Educ.* 13, 2587–2596. doi: 10.4236/ce.2022.138165

Li, J., Shi, Z., and Xue, E. (2020). The problems, needs and strategies of rural teacher development at deep poverty areas in China: rural schooling stakeholder perspectives. *Int. J. Educ. Res.* 99:101496. doi: 10.1016/j.ijer.2019.101496

Li, X., and Sun, X. (2018). Interface design and development of coating materials in lithium–sulfur batteries. *Adv. Funct. Mater.* 28:1801323. doi: 10.1002/adfm.201801323

Li, M., Zhu, J., and Wang, L. (2022). Classroom teaching performance evaluation model guided by big data and Mobile computing. *Wireless Commun. Mobile Comput.*:2022. doi: 10.1155/2022/2084423

Lin, S. S., Shen, S. L., Zhou, A., and Xu, Y. S. (2021). Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods. *Autom. Constr.* 122:103490. doi: 10.1016/j.autcon.2020.103490

Lu, H., Zhang, Y., Li, Y., Jiang, C., and Abbas, H. (2020). User-oriented virtual mobile network resource management for vehicle communications. *IEEE Trans. Intell. Transp. Syst.* 22, 3521–3532. doi: 10.1109/TITS.2020.2991766

Matos Pedro, E., Alves, H., and Leitão, J. (2020). In search of intangible connections: intellectual capital, performance and quality of life in higher education institutions. *High. Educ.* 83, 243–260. doi: 10.1007/s10734-020-00653-9

Ozanich, E., Gerstoft, P., and Niu, H. (2020). A feedforward neural network for direction-of-arrival estimation. *J. Acoust. Soc. Am.* 147, 2035–2048. doi: 10.1121/10.0000944

Renigier-Biłozor, M., Janowski, A., and d'Amato, M. (2019). Automated valuation model based on fuzzy and rough set theory for real estate market with insufficient source data. *Land Use Policy* 87:104021. doi: 10.1016/j.landusepol.2019.104021

Rodríguez-Campo, L., Alén-González, E., Antonio Fraiz-Brea, J., and Louredo-Lorenzo, M. (2022). A holistic understanding of the emotional experience of festival attendees. *Leis. Sci.* 44, 421–439. doi: 10.1080/01490400.2019.1597790

Runhong, L. (2022). Analysis on the construction of document resources in libraries of higher vocational colleges in central and Western China. *Libraly J.* 41:87. doi: 10.13663/j.cnki.lj.2022.01.011

Shen, C. W., and Ho, J. T. (2020). Technology-enhanced learning in higher education: a bibliometric analysis with latent semantic approach. *Comput. Hum. Behav.* 104:106177. doi: 10.1016/j.chb.2019.106177

Sun, X. (2020). Exploration and practice of "internet+ maker education" university innovative entrepreneurship education model from the perspective of positive psychology. *Front. Psychol.* 11:891. doi: 10.3389/fpsyg.2020.00891

Taufik, M. (2020). Strategic role of Islamic religious education in strengthening character education in the era of industrial revolution 4.0. *Jurnal Ilmiah Islam Futura* 20, 86–104. doi: 10.22373/jiif.v20i1.5797

Thamallah, A., Sakly, A., and M'Sahli, F. (2019). A new constrained PSO for fuzzy predictive control of quadruple-tank process. *Measurement* 136, 93–104. doi: 10.1016/j.measurement.2018.12.050

Wang, N., Lv, X., Sun, S., and Wang, Q. (2022). Research on the effect of government media and users' emotional experience based on LSTM deep neural network. *Neural Comput. Appl.* 34, 12505–12516. doi: 10.1007/s00521-021-06567-6

Xiong, G., Cao, X., Hamm, N. A., Lin, T., Zhang, G., and Chen, B. (2021). Unbalanced development characteristics and driving mechanisms of regional urban spatial form: a case study of Jiangsu Province, China. *Sustainability* 13:3121. doi: 10.3390/su13063121

Yao, Y., and Shekhar, D. K. (2021). State of the art review on model predictive control (MPC) in heating ventilation and air-conditioning (HVAC) field. *Build. Environ.* 200:107952. doi: 10.1016/j.buildenv.2021.107952

Yi, X., Li, G., Liu, Y., and Fang, F. (2021). Event-triggered H∞ filtering for nonlinear networked control systems *via* TS fuzzy model approach. *Neurocomputing* 448, 344–352. doi: 10.1016/j.neucom.2021.03.081

Zeng, H. B., Teo, K. L., He, Y., and Wang, W. (2019). Sampled-data stabilization of chaotic systems based on a TS fuzzy model. *Inf. Sci.* 483, 262–272. doi: 10.1016/j.ins.2019.01.046

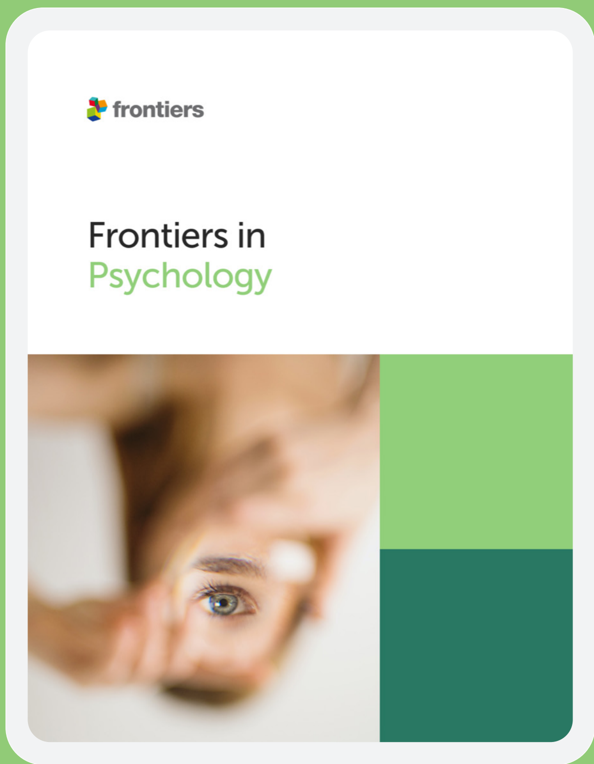
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact