

REVIEW

Open Access



Factors predicting mathematics achievement in PISA: a systematic review

Xiaofang Sarah Wang^{1*} , Laura B. Perry¹, Anabela Malpique² and Tobias Ide³

*Correspondence:
sarah.wang@murdoch.edu.au

¹ School of Education, Murdoch University, 90 South St, Murdoch, WA 6150, Australia

² School of Education, Edith Cowan University, 270 Joondalup Drive, Joondalup, WA 6027, Australia

³ Murdoch University, Harry Butler Institute, 90 South St, Murdoch, WA 6150, Australia

Abstract

The Programme for International Student Assessment (PISA) has become the world's largest comparative assessment of academic achievement. While hundreds of studies have examined the factors predicting student achievement in PISA, a comprehensive overview of the main predictors has yet to be completed. To address this gap, we conducted a systematic literature review of factors predicting mathematics performance in PISA. Guided by Bronfenbrenner's ecological model of human development, we synthesized the findings of 156 peer reviewed articles. The analysis identified 135 factors that fall into five broad categories: individual student, household context, school community, education systems and macro society. The analysis uncovered seven factors that are consistently associated with math achievement in PISA. Student grade level and overall family SES (socio-economic status) are consistently positively associated with math achievement while five factors are consistently negatively associated with math achievement: student absenteeism and lack of punctuality, school repeating and dropout rate, school prevalence of students' misbehavior, shortage of teachers and general staff, and student-centered instruction. Fourteen factors tend to be positively or negatively associated with math achievement. The explanatory power of many other factors, however, remain mixed. Explanations for this result include methodological differences, complex interactions across variables, and underlying patterns related to national-cultural context or other meso or macro-level variables. Implications for policy and research are discussed.

Keywords: Math, Academic performance, PISA, Secondary school, Ecological model

Introduction

Results of international large-scale assessments (ILSAs) in education have attracted increasing attention from policy makers, educational researchers and practitioners, and the public. Among all the ILSAs, the Programme for International Student Assessment (PISA) has become the largest internationally executed assessment (Martens & Niemann, 2013). PISA has been administered every three years since 2000 and the number of participating countries¹ has grown substantially, from 32 countries in

¹ OECD uses the term "countries/economies" in its reports because some regions which participated in PISA are not independent countries. For example, Macao (China), Hong Kong (China), Beijing-Shanghai-Jiangsu-Zhejiang (China) are three separate PISA economies. For the convenience of this article, the word "countries" is used to refer to all the participating regions in PISA.

2000 to 85 countries in the 2022 cycle (OECD, 2022), with 170 participating nations anticipated by 2030 (Xiaomin & Auld, 2020). PISA aims to evaluate how well 15-year-old students, who are approaching the end of compulsory schooling, can apply learned knowledge and skills to real-life situations (Hopfenbeck & Görge, 2017; OECD, 2022).

In each PISA cycle, three major cognitive domains—reading, mathematics and science—are assessed with one subject domain tested in detail. The focused domain in the 2022 cycle was mathematics. The impact of PISA cannot be ignored. It provides references to education system improvement and informs policy decisions across many countries (e.g., Baird et al., 2016; Rocher & Hastedt, 2020; Thompson et al., 2016).

International PISA-related research has also experienced dramatic growth in the past two decades (Addey et al., 2017; Hernández-Torrano & Courtney, 2021). Because “mathematics and science scores are highly invariant and can be used to compare countries” (Odell et al., 2021, p. 1), many researchers have investigated math achievement from a wide variety of perspectives across multiple countries. Some researchers focused on one or few factors in predicting math achievement in one or few countries (e.g., Sälzer & Heine, 2016). Researchers often focus on investigating one or few factors in multiple countries (e.g., Fang et al., 2013); many factors in one or a few countries (e.g., Perera & Asadullah, 2019); or many factors in many countries (e.g., Munir & Winter-Ebmer, 2018). Considering the various combinations of diverse factors and multiple countries, researchers have found it increasingly complex to provide a clear picture of the factors predicting math achievement across countries and assessment cycles. The OECD (Organisation for Economic Co-operation and Development) reports provide an overview of factors and single countries based on data from a single PISA cycle. However, factors are usually analyzed using one type of statistical analysis, which may leave unknown findings that would be uncovered from other approaches. The reports also typically use pooled data from all the participating countries, which may obscure cross-national differences. Country reports produced by the OECD also offer an overview of single countries but no direct comparison across countries is made. Finally, contradictory results have also been found between secondary analyses published as peer-reviewed journal articles and OECD reports (e.g., Agirdag & Vanlaar, 2018). We therefore agree with Gamazo and Martínez-Abad (2020, p. 12) that there is a clear need to “produce a thorough systematic review to explore the different methodologies employed, and evidences on the impact of diverse variables on student performance”.

As of yet, no systematic review has examined the predictors of PISA math achievement. Gutiérrez-de-Rozas et al. (2022) conducted a systematic review of 80 meta-analyses and presented the relationship between personal, family, school and teacher variables and academic achievement, but did not focus on math achievement or PISA performance. Another three reviews have been conducted about other aspects of PISA: Hopfenbeck et al. (2018) provided a broad overview of PISA-related research but did not examine math achievement; the scoping review of Odell et al. (2020) studied the relationship between Information and Communication Technology (ICT) and math/science in PISA; and Teig et al. (2022) conducted a systematic review on science teaching and learning using TIMSS and PISA data. A systematic review investigating major factors driving PISA math achievement hence remains a key research gap.

Addressing this knowledge gap, this study presents a comprehensive picture of the factors predicting PISA math performance. To do so, we proceed in five steps. First, we explain our theoretical framework before outlining our core method, a systematic literature review. We then present our findings and discuss them in the light of existing theoretical and empirical evidence. The limitations of our study and avenues for future research are summarized before the concluding section.

Theoretical framework

The theoretical framework guiding this study is Bronfenbrenner's ecological theory of human development (Bronfenbrenner, 1981), which emphasizes the importance of system-related factors and "the interaction of biological factors and the contexts in which people develop" (Rosa & Tudge, 2013, p. 251). Academic achievement is not solely due to the efforts of teachers or schools or individual students, and studies (e.g., Caro et al., 2016; Scheeren, 2022) have shown substantial interaction among various factors predicting math achievement. We therefore chose Bronfenbrenner's "person-process-context model" because it captures a wide range of factors and their complex interactions, as well as interactions with social contexts.

Method

The method and reporting of this review were guided by the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 (Page et al., 2021). The following steps were included: (1) search strategies were developed and applied to search in databases at the literature identification stage; (2) records were screened according to the eligibility criteria; (3) after confirming the included literature, data were extracted/synthesized and the quality of papers was appraised.

Literature identification

During the process of determining the search terms, we followed a comprehensive strategy to refine the search terms to ensure that all possible relevant papers were captured by the final search. A list of keywords "PISA", "math", "performance", "achievement", "result", "score", "outcome" was generated by the authors based on prior-knowledge, and these keywords were initially entered into a single database (the Education Resources Information Center-ERIC). The results of this scoping search were reviewed to identify additional keywords and synonyms. Two additional keywords "attainment" and "success" were added to the search terms. We also checked in-built thesauri in databases and subject headings that were assigned to relevant retrieved articles, but no additional synonyms were found. The search terms were also verified by intensive discussion between the authors. Afterwards, the refined search terms were formulated.

For our search strategy, we combined relevant keywords using Boolean operators AND, OR and "truncation *" to cover variations in the keywords. We created seven search strings (e.g., PISA OR "Program*" for International Student Assessment" AND performance* AND math*; see Additional file 1: Appendix S1 for the complete list of Boolean search strings used) and applied them to five bibliographic databases (APA PsycInfo, ERIC, Informit, Scopus and Web of Science), set to three major areas (title, abstract, and keywords). The electronic search was limited to peer-reviewed journal

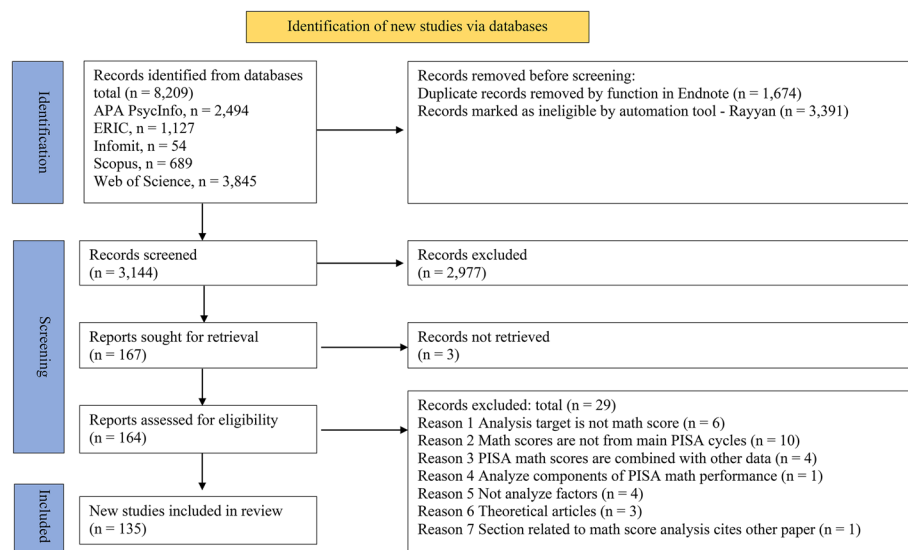


Fig. 1 PRISMA 2020 flow diagram for the systematic review

articles published in the English language between 1 September 2011 and 30 September 2021. The search yielded 8,209 records, which was reduced to 6,535 records after removing duplicate records (see Fig. 1 for detailed information).

The remaining 6535 records were copied to an automation tool—Rayyan (<http://rayyan.qcri.org>). Rayyan is “a process of semi-automation” with “a high level of usability” and it helps accelerate the initial screening of abstracts and titles (Ouzzani et al., 2016, p. 1). Certain keywords, such as patients, clinical, disorder, disease, were used to exclude articles. Some keywords, such as model, sex, stress, functioning, yielded results which could be relevant to the review question. Therefore, manually reviewing the article titles and abstracts when needed was performed to each record identified by Rayyan. A total of 3,391 records identified in Rayyan were removed due to their irrelevance to the study, leaving 3,144 articles to be manually screened.

Record screening

Eligibility criteria were discussed among all authors, pre-defined and applied to the primary screening of the titles and abstracts of 3144 articles. The studies were included if they met all the following criteria:

1. Peer-reviewed journal articles;
2. Published in English;
3. Published from 1 September 2011 to 30 September 2021;
4. Examined reasons/factors predicting PISA math performance;
5. Analyzed math scores from any of the PISA main cycles (2000, 2003, 2006, 2009, 2012, 2015, 2018).
6. Included data from other sources, but PISA math scores were analyzed separately. Conceptual or theoretical articles that do not analyze PISA math scores were excluded.

7. Published in journals which are ranked in the top quartile (Q1) in any discipline in Scimago Journal & Country Rank (<https://www.scimagojr.com>) at least once between year 2016 and year 2020.

Publication's ranking (criterion 7) was not applied initially in the inclusion criterion. However, screening the titles and abstracts without including this ranking criterion yielded over 500 results, which would have constrained subsequent screening and analysis. Q1 journals are also known to have particularly strict quality controls. After the articles were restricted to those published in Q1 journals, the results were reduced to 167 articles (see Fig. 1). Considering the ranking fluctuation of some journals over the years, the condition was set to any journal which appeared in Q1 in Scimago Journal & Country Rank between 2016 and 2020 at least once, to reduce potential bias related to the ranking. 2021 ranking was not considered because it was not available at the beginning of this review.

All 167 papers were retained, except for three articles which were written in languages other than English, even though their titles and abstracts were available in English. This left 164 papers for the second screening of the full text. Twenty-nine articles did not meet the inclusion criteria and were removed after detailed discussion among the authors. A total of 135 articles were eligible from the first run of databases.

Before final data synthesis, and considering the length of the review being taken, we conducted a second screening limiting the timeframe from 31 August 2021 to 1 July 2022. We applied the same search strategies in the same five databases, which yielded 1169 records in total. After the process of removing the duplicates and screening the title and abstract according to the inclusion criteria shown above, 26 records remained. The full text of all the articles were retrieved and five articles were excluded according to the inclusion criteria, leading to an additional 21 articles, for a total of 156 articles included in this systematic review.

Data extraction and synthesis

All 156 eligible articles were imported to NVivo for data extraction and synthesis (see Additional file 2: Appendix S2 for the summary of reviewed articles). During coding, an inductive approach was initially applied (Braun & Clarke, 2006, 2021). Codes and themes were formulated when the articles were examined. After the data from 50 papers were extracted, an overview of the emerging themes was set. Both deductive and inductive approaches were then utilized in examining the remaining papers as new themes occasionally appeared. Data were coded into fine-grained themes which were recorded separately. Since emerging patterns echoed Bronfenbrenner's theory of human development, a synthesis framework was developed based on the phase two "person-process-context model" (Bronfenbrenner & Morris, 1998, 2006; Rosa & Tudge, 2013), as shown in Fig. 2.

Interrater reliability

Three interrater reliability (IRR) tests were conducted at primary screening, second screening and data extraction to reduce human factors that can affect decision-making in a systematic review. We adopted the most common consensus estimate of IRR, namely the percent agreement statistic (Salkind, 2011). During the primary screening

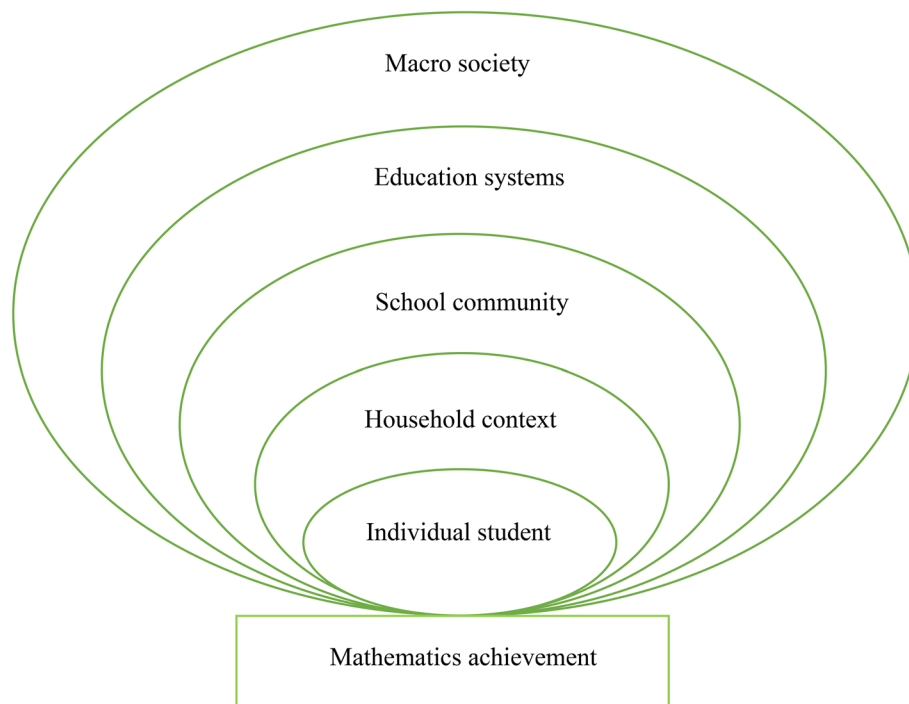


Fig. 2 synthesis framework Source: authors' own compilation based on Bronfenbrenner's ecological theory of human development

of titles and abstracts, a random selection of 10% of the 3144 articles were assigned to the second and third authors. Their screening results were compared to the results of the first author. Agreement was not reached for a total of 40 records and therefore these were assigned to the fourth author, following which 32 records were resolved. The final eight records were discussed among all the authors until consensus was reached. The IRR agreement was high (94%).

At the second screening, 10% of 164 records were assigned to the second and third authors to read the full texts. The screening results of the second and third authors were compared to the results of the first author, which yielded five records in disagreement. Further discussion was made among these three coders until consensus was reached. The IRR agreement rate was 85%. Both IRR rates have exceeded the commonly acceptable percentage of 80% (Belur, et al., 2021).

The third layer of IRR check occurred during data extraction. The first author selected the articles thought not to clearly meet inclusion and exclusion criteria ($n = 16$). Subsequently, the second and third authors read the articles in detail and discrepancies were discussed and resolved.

Risk of bias assessment

The articles' quality was also assessed against the Johns Hopkins nursing evidence-based practice rating scale (Dang et al., 2021), which comprises two main aspects—evidence level and quality ratings. The three levels of evidence are true experimental designs (Level I), quasi-experimental studies (Level II) and nonexperimental designs (Level III). Quality rating is classified into three categories: high quality (A), good quality (B), and

low quality (C). It includes considerations as to whether studies provide sufficient sample size, generalizable results, adequate control, and definitive conclusions, to name a few (Dang et al., 2021). All included studies in this review were Level III non-experimental studies but the quality of papers ranged from high quality A to low quality C. Limitations in quality generally related to how much control was applied to variables, how definitive the conclusions were, and how generalizable the results were. The quality rating for each article is included in Additional file 2: Appendix S2. No article was excluded due to low quality and all 156 articles were synthesized.

Results

In this article, we reviewed 156 studies that investigated factors predicting mathematics achievement in PISA. Our aim was to capture as many of the factors that have been investigated as possible and identify the major patterns and key factors emerging from research in the field. In the following section, we will first provide an overview of included studies, followed by a detailed explanation of each major factor.

Overview of included studies

One-third of the studies ($n=55$) investigated math achievement in single countries. Twenty-two countries were examined, with 20% of the studies examining data from the US, followed by Spain, Italy, Australia and Turkey. Another dominant proportion of studies (30%, $n=47$) investigated math performance in more than 30 countries. Some used pooled data from multiple countries (e.g., Borgonovi & Ferrara, 2020); some compared individual countries (e.g., Caro et al., 2016); and others combined both approaches in one study (e.g., Bhutoria & Aljabri, 2022). Approximately 65% of the reviewed studies used the PISA 2012 cycle, which is not surprising as PISA 2012 is the latest cycle with available data focused on the mathematics domain. Less than 20% of the studies ($n=30$) examined more than one PISA cycle.

Regarding the methodological approaches, almost all studies used a form of correlational analyses, with hierarchical linear regressions (30%) being the most common. Only two studies used other methods: Borgna, (2016), who used Qualitative Comparative Analysis (QCA), Lee and Borgonovi (2022), who used Necessary Condition Analysis (NCA). Two common limitations were identified among the reviewed studies. First, factors included in each study varied and variables were omitted due to different reasons. Omitted factors may explain variation in math achievement (e.g., Leung, 2014; Sousa et al., 2012). Second, most studies stated that no causal claims can be made due to the cross-sectional nature of PISA data (e.g., Lezhnina & Kismihók, 2022).

Emerging categories

We used Bronfenbrenner's theory of human development phase two "person-process-context model" to categorize factors into five levels, namely individual student, household context, school community, education systems, and macro society. One hundred and thirty-five factors were identified (see Additional file 3: Appendix S3). Due to the large number of studies included in this review, we only report factors in the main text below if five or more studies investigated their association with mathematics achievement, which leads to 57 major factors. Figure 3 shows the number of

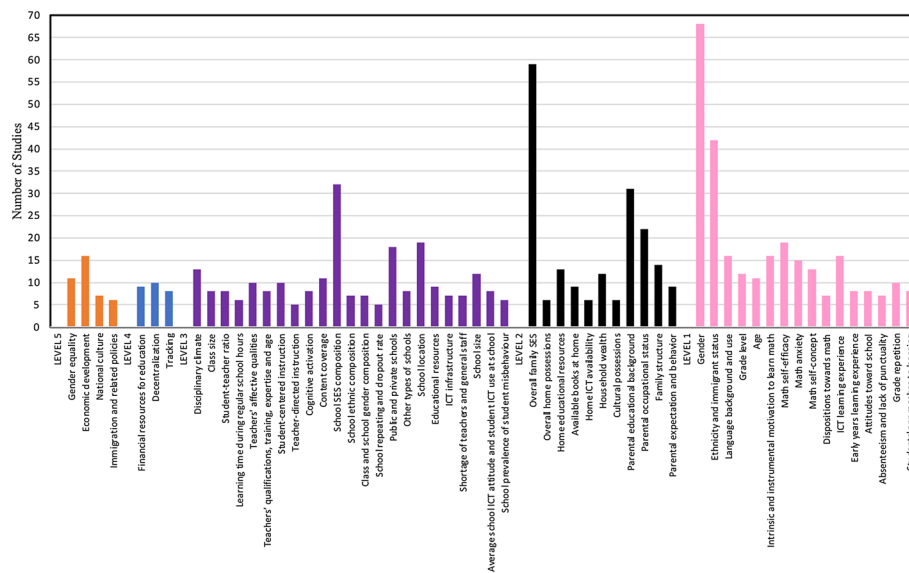


Fig. 3 number of relevant studies of major factors

relevant studies that examined each of these 57 major factors. A detailed map of factors associated with each article can be found in Additional file 2: Appendix S2. The definition of each factor referenced in this review can be found in Additional file 3: Appendix S3. It should be noted that discrepancies existed among studies in terms of how to measure some factors. Even when studies use the definition of PISA ESCS (economic, social and cultural status) as a measure of family socio-economic status (SES), “the operational definition of ESCS has changed in almost every cycle” (Avviasi, 2020, p. 30). Our definition incorporates aspects from all relevant studies in the review.

Gender (68 relevant studies) was the most frequently investigated factor, followed by overall family SES. However, if considering the sub-factors under family SES, such as parental educational background, 93 studies in this review investigated family SES (60% of reviewed papers). The other two factors attracting extensive research were ethnicity and immigrant status, and school SES composition. Factors at Level 4 and 5 were not as extensively investigated as factors at the other three levels.

Figure 4 provides an overview of the association between these major factors and math achievement. Specifically, we indicate whether the relationship is positive, negative, insignificant (as a measure of statistical significance), or mixed (meaning that the association differed across or within studies). Among the 57 major factors, two major factors identified in the review were consistently positively associated with math achievement, namely student grade level and overall family SES. Five factors were consistently negatively associated with math, namely student absenteeism and lack of punctuality; school repeating and dropout rate; school prevalence of students’ misbehavior; shortage of teachers and general staff; and student-centered instruction. Twelve major factors tended to be positively associated with math and two major factors tended to be negative, meaning such association were reported in 80% or more of the relevant studies. For the remaining 36 major factors, we found mixed results.

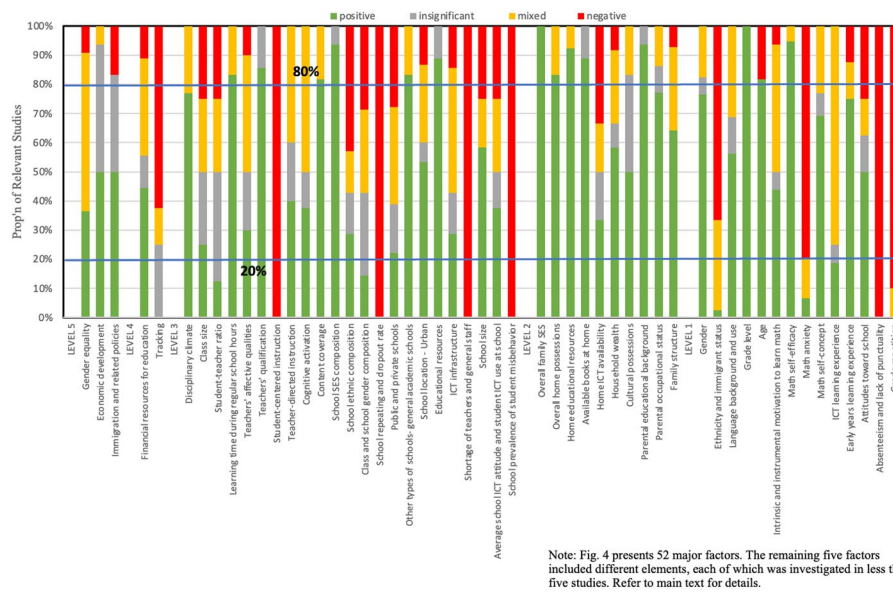


Fig. 4 overview of the association between major factors and math achievement

The main text below is divided into five sections with each section corresponding to major factors at each level. We introduce each section with a brief summary, followed by detailed results related to major factors at that level.

Level 1: individual student factors

Six major factors emerged at Level 1, namely individual/demographic characteristics; psychological factors; learning opportunities; behavior and engagement at school; learning outcomes; students’ approaches to learning. The definition of these factors can be found in Additional file 3: Appendix S3. Student grade level was consistently positively associated with math achievement while absenteeism/lack of punctuality was consistently negatively associated with math. Student age and math self-efficacy tended to be positively associated with math achievement, while math anxiety and grade repetition tended to be negatively associated with math achievement. Results related to all the other major factors, such as gender, ethnicity/immigrant status, were mixed within and across studies.

Individual/demographic characteristics

Individual/demographic characteristics included four major sub-factors, namely gender; ethnicity and immigrant status; grade level; and age. The key sub-factor under ethnicity and immigrant status was language background and use.

Gender Our review found 68 relevant studies reporting mixed results regarding the relationship between gender and math achievement. Fifty-two studies found that gender was associated with math achievement favoring male students (e.g., Gevrek et al., 2020; Hána et al., 2017). Four single-country studies found an insignificant association of gender with math achievement, such as in Portugal (Melkonian et al., 2019), and in eight Western countries (Hoffmann, 2022). Twelve studies found mixed results within studies,

seven of which reporting associations between gender and math performance depending on countries (e.g., Kim & Law, 2012; Zhang et al., 2022). Three studies found a positive association between male and math achievement using overall/pooled data, but mixed results were reported in individual countries (Sortkær & Reimer, 2018), in the top grade percentile (Matějů & Smith, 2015), and among different racial groups (Schmidt et al., 2021). Females were found to outperform male students in Georgia (Kameshwara et al., 2020), Malaysia (Thien & Ong, 2015), among first generation immigrants in the US (Kim, 2018), white females in the US (Schmidt et al., 2021), and on the high-reading-demand math questions, regardless of their reading literacy level (Ajello et al., 2018).

Ethnicity and immigrant status Our review found 42 relevant studies reporting mixed results regarding the relationship between ethnicity/immigrant status and math achievement. Twenty-eight studies found ethnicity/immigrant status was negatively associated with lower math scores, with the performance of immigrants lower than native students (e.g., Luschei & Jeong, 2021; Radl et al., 2017). Compared to second-generation students and native students, first-generation students were found to score the lowest (e.g., Rodríguez et al., 2020; Spörlein & Schlueter, 2018). Ferraro (2018) found the opposite result, with native students negatively associated with lower math scores in Italy, although the major focus of the study was on ICT. Thirteen studies found mixed results within studies (e.g., Martin et al., 2012). For instance, immigrant status was not significantly associated with math in all countries investigated (e.g., Cheung, 2017; Cobb-Clark et al., 2012).

Relating to the impact of race, eight studies investigated the various races in the US. Seven studies found that African and Hispanic students scored lower than White students while Asian ethnicity was positively associated with higher math results (e.g., Li et al., 2015; Pivovarova & Powers, 2019b). However, Cheng et al. (2014) combined the racial groups with generations of migration and found substantial difference across racial groups, with one example of the difference between the first-generation African/Asian immigrants and native students being statistically insignificant. The factors included in these eight studies differed across studies, but Cheng et al. (2014) was the only study which used t-test and simple linear regression analysis. Another three studies investigated the impact of race in two other countries. Indigenous Australians were found to underperform compared to non-Indigenous students (Dockery et al., 2020; Posso, 2016). Jewish students were found to perform higher than Arab students in Israel (Razer et al., 2018).

Relating to the age of arrival in the destination country for the immigrant students, six out of seven studies found it to be negatively associated with math achievement, with a strong negative association among immigrants who arrived around or after the age of 12 (e.g., Borgonovi & Ferrara, 2020; Isphording et al., 2016). Compared to these cross-national studies, Cheng et al. (2014) using data from the US found mixed results with respect to the age of students' arrival, depending on students' race.

Relating to the other factors which interacted with ethnicity and immigrant status, family background (e.g., family SES, household structure) and school background (e.g., content coverage) were found to reduce the negative impact of ethnicity and immigrant status on math achievement (e.g., Schmidt et al., 2021; Spörlein & Schlueter, 2018). Two studies found that the negative association between immigrant status and math

achievement became insignificant after accounting for school mean SES (Karakolidis et al., 2016) and students background—gender, race, language gap, parental education and wealth (Pivovarova & Powers, 2019b). Students' country of origin being highly-ranked in math was also found to reduce the negative impact of ethnicity/immigrant status (Giannelli & Rapallini, 2016).

Language background and use

Apart from the investigations of the impact of ethnicity/immigrant status in general, 16 studies investigated the impact of students' language background. Nine studies found that students' not speaking the test language/host country language at home was associated with lower math achievement, using multiple PISA cycles in single countries (e.g., Pivovarova & Powers, 2019b) and large cross-national studies (e.g., Orón Semper et al., 2021). Two studies found students' language background and use was insignificantly associated with math in Portugal (Melkonian et al., 2019), and Italy and Spain (Azzolini et al., 2012). The remaining five studies found mixed results within studies, three of which found speaking the test language at home positively associated with higher math achievement in some countries but not in others (Agirdag & Vanlaar, 2018; Borgna, 2016; Zhu & Kaiser, 2020). Two studies found mixed results within a single country. In Australia, speaking English as a second language was found to be negatively associated with math for immigrants in general but negligible for second-generation immigrants (Dockery et al., 2020). In Spain, the impact of test and home languages was found to vary according to regions and PISA cycles (Lopez-Agudo et al., 2021).

Grade level All 12 studies found a positive association between grade level and math performance, meaning the higher grade levels students were at, the higher their math scores (e.g., Aguayo-Téllez & Martínez-Rodríguez, 2020; Spörlein & Schlueter, 2018). This is not surprising, as students at higher grade levels have typically been exposed to more extensive math content, and this review has shown that content coverage is mostly positively associated with math achievement (e.g., Barnard-Brak et al., 2018). It is worth noting that while PISA is administered to 15-year-old students, rather than a particular grade level, most 15-year-old students in any given country are typically in a particular grade, although there may be some variation.

Age Age tended to be positively associated with math achievement in 11 relevant studies. Nine studies found a positive association between age and math achievement, meaning the older students were, the higher the math scores that they obtained (e.g., Giannelli & Rapallini, 2016; Rodríguez et al., 2020). Two studies found opposite results, with older students performing significantly worse in Australia (Posso, 2016) and in 62 countries (Schmidt et al., 2015). Closely examining two studies which investigated the impact of age in Australia but found contradictor results, both studies used regression models but Dockery et al. (2020) used PISA 2015 data while Posso (2016) used PISA 2012 data. The factors included in both studies also differed.

Psychological factors

Psychological factors included two major sub-factors, namely drive and motivation; and math self-beliefs, dispositions and participation in math-related activities. The key sub-factor under drive and motivation was intrinsic and instrumental motivation to learn math. The key sub-factors under math self-beliefs, dispositions and participation in math-related activities were math self-efficacy, math anxiety, math self-concept, and dispositions towards math.

Intrinsic and instrumental motivation to learn math Our review found 16 relevant studies reporting mixed results regarding the relationship between intrinsic/instrumental motivation and math achievement. Seven studies found a positive association between motivation and math achievement, using data from a single country like Australia (Gabriel et al., 2020) and Israel (Razer et al., 2018) or large cross-national samples (e.g., Bhutoria & Aljabri, 2022). Another seven studies found mixed results in a limited number of countries (e.g., Liu et al., 2022; Zhang et al., 2022). We were unable to uncover clear patterns that could explain the mixed results, although three studies found the impact of motivation varied according to the analytical methods used (Chen & Lin, 2020) or the interaction with other factors, such as math anxiety (Xiao & Sun, 2021) or attitudes toward school (Pitsia et al., 2017). The remaining two studies used different PISA cycle data in Turkey and found contradictory results, namely negative (Niehues et al., 2020) and insignificant (Yıldırım, 2012) associations between student motivation and math performance. Although the measure of motivation in both studies seemed alike, the factors included in the studies and the analytical methods differed.

Math self-efficacy Math self-efficacy tended to be positively associated with math achievement in 19 relevant studies. Eighteen studies found self-efficacy positively associated with math achievement (e.g., Brow, 2019; Zhao & Ding, 2019), with only one study, a three-country comparison, reporting mixed results (Thien et al., 2015). In terms of interaction with other factors on math achievement, self-efficacy was found to positively interact with disciplinary climate (Cheema & Kitsantas, 2014); perseverance, math intentions, and work ethic (Kitsantas et al., 2020); family SES (Niehues et al., 2020); and opportunity to learn (OTL; F. Wang, Wang et al., 2022).

Math anxiety Math anxiety tended to be negatively associated with math achievement in 15 relevant studies. Twelve studies found math anxiety was negatively associated with math achievement, with 10 studies using data from a single or small number of countries (e.g., Fan et al., 2019; Zhao & Ding, 2019) and two large cross-national samples (Lee & Stankov, 2013, 2018). Two small-n comparative studies (Cheung, 2017; Thien et al., 2015) found mixed results across several Asian countries. In terms of interaction with other factors, math anxiety was found to be negatively associated with math self-concept (Pitsia et al., 2017).

Math self-concept Our review found 13 relevant studies reporting mixed results regarding the relationship between math self-concept and math achievement. Nine studies found a positive association between math self-concept and math achieve-

ment, with eight studies using data from a single or small number of countries (e.g., Guglielmi & Brekke, 2018; Parker et al., 2014) and one study using data from 41 countries (Lee & Stankov, 2013). Four studies (Cheung, 2017; Thien & Ong, 2015; Thien et al., 2015; Zhang et al., 2022) found mixed results across several Asian countries, and even within the same country (Singapore).

Dispositions towards math Our review found mixed results regarding the relationship between dispositions towards math and math achievement. Dispositions towards math contained math intentions and subjective norms in math. Four studies investigated the impact of subjective norms in math, three of which found a negative association between subjective norms in math and math achievement, meaning students who perceived a higher level of friends/parents enjoying and valuing mathematics tended to have lower math scores in the US, China and Qatar (e.g., Areepattamannil et al., 2016). Munir and Winter-Ebmer (2018) found a positive association, with higher subjective norms reducing male–female gaps using data from 65 countries. In terms of math intentions investigated in three relevant studies, two studies found mixed results depending on countries, namely negative in the US but positive in East Asia and Mexico (Guglielmi & Brekke, 2018); and negative in Romania, positive in Finland and Australia, and insignificant in Singapore (Zhang et al., 2022). Kitsantas et al. (2020) found a positive association, with high math intentions positively associated with math outcome in the US. Despite using the same PISA cycle and sample size, the two studies investigating the impact of math intentions in the US produced contradictory results, but the analytical methods and factors included in each study varied (Guglielmi & Brekke, 2018; Kitsantas et al., 2020).

Learning opportunities

Learning opportunities included two major sub-factors, namely ICT learning experience; and early years learning experience.

ICT learning experience Our review found 16 relevant studies reporting mixed results regarding the relationship between ICT learning experience and math achievement. Twelve studies found mixed results related to student ICT, with ten studies examining a single to small number of countries (e.g., Meng et al., 2019; Ünal et al., 2022) and two large cross-national studies using different PISA cycles (Bhutoria & Aljabri, 2022; Hu, Gong et al., 2018). Among these 12 studies, the association between ICT use/attitude and math achievement varied substantially cross-nationally, ranging from positive to negative to insignificant, with clear patterns difficult to identify. Three studies found a positive association between math achievement and ICT accessibility using five PISA cycles in 48 countries (Erdogdu, 2022); between math achievement and spending more time in educational ICT use and leisure ICT use in 39 countries (Skryabin et al., 2015); and between math and more favorable attitude towards ICT in Spain (Tourón et al., 2019). The remaining study investigated the impact of one type of leisure ICT use (videogaming) in 22 OECD countries and found an insignificant association with math achievement (Drummond & Sauer, 2014).

Five studies investigated the impact of the intensity of ICT use, four of which found the moderate use of ICT was more positively associated with higher math scores compared to no or excessive use (e.g., Bhutoria & Aljabri, 2022). For example, students who reported some Internet use outside of school were found to score higher than students who reported no use or 6+ hours use per day (Rozgonjuk et al., 2021). The intensity of general ICT use and the earlier age students first use ICT were found to be negatively associated with math achievement (Navarro-Martinez & Peña-Acuña, 2022), as was the higher ICT use for social networks (Posso, 2016).

Early years learning experience Our review found eight relevant studies reporting mixed results regarding the relationship between early years learning experience and math achievement. Six studies found a positive association between early years learning experience and math achievement, meaning having attended pre-school contributed to higher math scores, using three PISA cycles in a small number of countries (e.g., Giannelli & Rapallini, 2019) and one large cross-national study (Gamazo & Martínez-Abad, 2020). A negative association was found in Mexico by Aguayo-Téllez and Martínez-Rodríguez, (2020). The remaining study, a comparative study of five East Asian contexts, found positive association in two countries and an insignificant association in three countries (Cheung, 2017).

Behavior and engagement at school

Attitudes toward school Our review found eight relevant studies reporting mixed results regarding the relationship between attitudes toward school and math achievement. Four studies found that positive attitudes toward school were positively associated with math achievement, using data in India (Areepattamannil, 2014) and the US (Gjicali & Lipnevich, 2021), large cross-national samples (e.g., King et al., 2020). Two studies found a negative association between attitudes toward school and math, in Greece (Pitsia et al., 2017) and in 41 countries using PISA 2003 data (Lee & Stankov, 2013). Lee (2016) found an overall insignificant direct and indirect association between attitudes toward school and math using PISA 2003 and 2012 data in 41 countries, with some countries showing a positive relationship and others a negative relationship. Similar cross-national variation was found by Thien et al. (2015) in their comparative study of three southeast Asian countries.

Absenteeism and lack of punctuality All seven studies found a negative association between absenteeism/lack of punctuality and math achievement, with four studies using data from single countries (e.g., Fernández-Gutiérrez et al., 2020; Sälzer & Heine, 2016) and three large cross-national studies (e.g., Yamamura, 2019).

Learning outcomes

Grade repetition Grade repetition tended to be negatively associated with math achievement in 10 relevant studies. Nine studies found a negative association between grade repetition and math achievement, with seven studies using data from single to a limited number of countries (e.g., Salas-Velasco et al., 2021) and two large cross-national samples (Lee & Stankov, 2018; Luschei & Jeong, 2021). One small-n comparative study of five East

Asian countries (Cheung, 2017) found a negative association in four countries, and an insignificant association in the fifth.

Students' approaches to learning

Our review found mixed results regarding the relationship between students' approaches to learning and math achievement, depending on types of approaches, as well as within approaches. Four out of eight studies investigated the impact of meta-cognitive strategies and all of them found it positively associated with math achievement, using data from Asian countries (e.g., Wu et al., 2020) and one large cross-national sample (Bhutoria & Aljabri, 2022). Four studies investigated the impact of memorization and all of them found it negatively associated with math performance, using data from a single or small number of countries (e.g., Sousa et al., 2012) and one large cross-national sample (Lee & Stankov, 2013). Three studies investigated the impact of elaboration, showing mixed results. One of these studies found a positive relationship in India (Areepattammannil, 2014); another study found positive, negative and insignificant results that varied cross-nationally in six East Asian contexts (Wu et al., 2020); and a third study using data from 41 countries found a negative association (Lee & Stankov, 2013). Two studies investigated the impact of competitive learning and cooperative learning, finding it to be negatively associated with math achievement using data from 41 countries (Lee & Stankov, 2013), but positively in Hong Kong, Japan, Korea and the US (except for insignificance with respect to cooperative learning in the US) (Ma & Ma, 2014).

Level 2: household context

Three major factors emerged at Level 2, namely family SES; family structure; and parental expectation and behavior. Family SES also contained three sub-factors. The definition of these factors can be found in Additional file 3: Appendix S3. The direct impact of family SES was consistently positively associated with math achievement. The overall impact of home possessions, home educational resources, available books at home and parental educational background tended to be positively associated with math achievement. Results related to all the other major factors were mixed within and across studies, such as household wealth and family structure.

Family SES

Fifty-nine out of 93 studies investigated family SES using the overall concept of this factor, while the other studies investigated three sub-factors. Both direct and indirect effects of overall family SES were examined. All 59 studies found a positive association between the direct effect of the overall family SES and math achievement (e.g., Sulis et al., 2020; Xiao & Sun, 2021). Family SES was found to have less impact compared to school SES (Liu et al., 2022; Spagnolo et al., 2020).

The indirect effect of SES on other factors impacting on math was also investigated. These factors which had a positive association were cognitive activation (Caro et al., 2016), general attitude towards school (Lee, 2016), disciplinary climate/self-efficacy/OTL (F. Wang, Liu, & Leung, 2022; F. Wang, Wang et al., 2022; Schmidt et al., 2015), gender (stronger for girls) (Zhu et al., 2018), and parents' education-related beliefs

hence leading to student self-concept and self-efficacy (Niehues et al., 2020). On the other hand, the results related to the interaction between family SES, math achievement and other factors were mixed within studies depending on the country's context (Caro et al., 2016; Hwang et al., 2018). For instance, Caro et al. (2016) found the interaction of teacher-directed instruction and SES was negative in some countries, but positive in others. Apart from the effect of the overall family SES, three sub-factors under family SES were also investigated, namely home possessions, parental educational background, and parental occupational status.

Home possessions Six out of 36 studies investigated home possessions using the overall concept of this factor, while the other studies investigated three sub-factors, namely home educational resources, household wealth, and cultural possessions. The overall home possessions tended to be positively associated with math achievement in six relevant studies. Five studies found overall home possessions were positively associated with math achievement (e.g., Lee & Stankov, 2018; Shapira, 2012). Azzolini et al. (2012) found mixed results relating to home possession, positive for first-generation students and insignificant for second-generation students in Spain and Italy. However, caution should be taken in interpretation due to the small sample size for the second-generation students in this study.

Home educational resources

Home educational resources tended to be positively associated with math achievement in 13 relevant studies. Twelve studies found the overall concept of home educational resources was positively associated with math achievement (e.g., Dockery et al., 2020; Tan, 2017), although it failed to be a necessary condition when Lee and Borgonovi (2022) applied NCA analytical method (but positive correlation). Tsai et al. (2017) found mixed results, with a strong relationship in East Asian countries (Taiwan, Japan, South Korea), but not in Western countries (the US, Germany, the Czech Republic).

When examining two smaller factors under home educational resources, eight out of nine studies identified a positive association of books at home with math achievement (e.g., Brow, 2019; Cordero & Gil-Izquierdo, 2018). On the other hand, three out of six studies found a negative or insignificant association between home ICT availability and math achievement (Fernández-Gutiérrez et al., 2020; Hu, Gong et al., 2018; Tan & Hew, 2019), while two studies found a positive association (Giannelli & Rapallini, 2016; Kim, 2018). Moreover, one study found that the association between home ICT availability and math achievement in Taiwan changed from a weak positive association to negative after other factors (ICT use pattern and family SES) were controlled for (Chiu, 2020). Mixed findings with this factor may be due to variations in the countries, student groups, and other variables included in each study.

Household wealth

Our review found 12 relevant studies reporting mixed results regarding the relationship between household wealth and math achievement. Seven studies found that household wealth had a positive association with math achievement (e.g., Marks & Pokropek, 2019; Pivovarova & Powers, 2019b), although it failed to be a necessary condition (although

still a positive correlation) when Lee and Borgonovi (2022) applied NCA analytical method. Posso (2016), however, found a negative association between household wealth and math achievement in Australia, meaning students from wealthier families scored lower in math. Kim and Law (2012) did not find household wealth moderating the effect between gender and math in South Korea and Hong Kong. Three studies found mixed results, depending on countries (Xie & Ma, 2019), the interaction with other factors (Barnard-Brak et al., 2018; Sousa et al., 2012), or different PISA cycles (Sousa et al., 2012).

In terms of the interaction with other factors, household wealth was found to be positively associated with math achievement in the US (Barnard-Brak et al., 2018; Pivovarova & Powers, 2019a, 2019b), but Barnard-Brak et al. (2018) found that the significant association of household wealth became statistically non-significant after accounting for OTL. Similarly, a negative association of the number of cars/computers at home with math was found to change to insignificant association after school factors were accounted for (Sousa et al., 2012).

Cultural possessions

Our review found six relevant studies reporting mixed results regarding the relationship between cultural possessions and math achievement. Three studies found that cultural possessions had a positive association with math achievement in large cross-national studies (Chiu, 2015; Lee et al., 2019; Tan & Hew, 2019). Two studies found insignificant association between cultural possessions and math achievement (Kim & Law, 2012; Lee & Borgonovi, 2022). Xie and Ma (2019) found mixed results, with the positive mediating effects of cultural possessions on the relationship between other factors (such as parental occupational status, household wealth) and math achievement in most countries but insignificant in few, such as Chile. Mixed results may be due to methodological differences, as suggested by the conflicting findings of Lee and Borgonovi (2022) and Lee et al. (2019). Both studies used the same PISA cycle for sixty or more countries, but the former removed five Asian contexts and added national development status as an additional predictor variable.

Parental educational background Parental educational background tended to be positively associated with math achievement in 31 relevant studies. Twenty-nine studies reported a positive association between parental educational background and math achievement, especially for parents with tertiary education qualifications (e.g., Fernández-Gutiérrez et al., 2020; Gimenez et al., 2018). Tertiary education was found to be a necessary condition for high math achievement (Lee & Borgonovi, 2022). Parental years of schooling was found to have a better predictive power compared to parental educational levels (Giannelli & Rapallini, 2019; Lee et al., 2019). Moreover, paternal educational level was found to be more important compared to maternal (Dockery et al., 2020; Tao & Michalopoulos, 2018). The impact of parental education was also found to be stronger for first-generation immigrants or immigrants in general (Azzolini et al., 2012; Shapira, 2012). Just two studies found an insignificant association between parental educational background and math achievement: Cobb-Clark et al. (2012), who found an insignificant association for late immigrant students in 34 OECD countries, and Tsai et al.

(2017), who found an insignificant association in six Western and East Asian countries after controlling for the factor of scholarly culture at home.

Parental occupational status Our review found 22 relevant studies reporting mixed results regarding the relationship between parental occupational status and math achievement. Seventeen studies found a positive association between parental occupational status and math achievement (e.g., Gimenez et al., 2018; Tsai et al., 2017), two studies found an insignificant association (Sousa et al., 2012; Stoet et al., 2016), and three studies (Erdogdu, 2022; Giannelli & Rapallini, 2016; Lee & Borgonovi, 2022) found mixed results that varied for paternal and maternal occupational status. Erdogdu's (2022) analysis of aggregated data from five PISA cycles in 48 countries found a positive association between math achievement and maternal occupational status but insignificant association for paternal occupational status. By contrast, Lee and Borgonovi's (2022) analysis of PISA 2012 data from 60 countries found that paternal occupational status was positively associated with math achievement and was a necessary condition, but maternal occupational status was neither significant nor necessary. The other study found that paternal part-time employment was associated negatively with math achievement, whereas maternal part-time employment showed a positive association (Giannelli & Rapallini, 2016).

Family structure

Our review found 14 relevant studies reporting mixed results regarding the relationship between family structure and math achievement. Nine studies found a positive association between two-parent families and math achievement (e.g., Cordero & Gil-Izquierdo, 2018; Sousa et al., 2012), with students living in extended households scoring lower math scores (Bokhove & Hampden-Thompson, 2022; Hillier et al., 2021; Yamamura, 2019). The remaining five studies reported mixed results, with associations that varied cross-nationally or by student group. For example, Azzolini et al. (2012) found the association between two-parent household structure and math achievement was positive for first-generation immigrant students but insignificant for second-generation students in Italy and Spain. Cheung (2017) found the association between two-parent household and math achievement was positive in one East Asian context but insignificant in another four. Radl et al. (2017) found that the absence of the father in a family had an insignificant association with math achievement in four out of 33 OECD countries. Dronkers et al. (2017) also found the absence of the father to be insignificant once the factor of individual truancy was controlled for, suggesting that paternal absence had an indirect negative association with math performance because it was associated with increased student truancy. Finally, a study conducted in India by Areepattamannil (2014) found a completely contradictory result from the rest of the literature, namely that students living in single-parent families had significantly higher scores in math in two states.

Parental expectation and behavior

Parental expectation and behavior included many sub-factors, and nine out of 10 studies found some factors to be significantly positively associated with math achievement.

Some of these examples were parents' higher educational, career, academic expectations (Hillier et al., 2021; Tan, 2015, 2017); parents considering math important (Areepattammannil et al., 2015); parental support, including emotional support (Karakus et al., 2022); visits to museums three to four times per year (Hillier et al., 2021); and knowing a child's teacher well (Hillier et al., 2021). However, parents talking about school experiences and spending time doing fun activities with children was found to be insignificant in predicting math success in Canada (Hillier et al., 2021). Mixed results were also found within and across studies. The association of parental pressure on schools' academic standards with math achievement was positive in three countries using PISA 2012 data (Perera & Asadullah, 2019), positive in ten OECD countries using PISA 2009 (Sousa et al., 2012), and insignificant in ten OECD countries using PISA 2006 (Sousa et al., 2012).

Level 3: school community factors

Two major categories emerged at Level 3, namely within-classroom factors, and school characteristics.

Within-classroom factors

Four major factors emerged at Level 3 within classrooms, namely classroom characteristics; teacher characteristics; pedagogy and assessment; and content coverage. Student-centered instruction was consistently negatively associated with math achievement. Learning time during regular school hours, teachers' qualification, and content coverage tended to be positively associated with math achievement. Results related to all the other major factors, such as student–teacher ratio and teachers' affective qualities, were mixed within and across studies.

Classroom characteristics Classroom characteristics contained four sub-factors, namely disciplinary climate; class size; student–teacher ratio; and learning time during regular school hours.

Our review found 13 relevant studies reporting mixed results regarding the relationship between disciplinary climate and math achievement. Ten studies found a positive association between disciplinary climate and math achievement using data from four PISA cycles in single countries (e.g., Cheema & Kitsantas, 2014) to multiple countries (Santibañez & Fagioli, 2016). Three cross-national studies, however, found mixed results within their studies depending on the countries investigated (e.g., Caro et al., 2016; Sortkær & Reimer, 2018). No negative association between classroom environment and math achievement was found.

Our review found eight relevant studies reporting mixed results regarding the relationship between class size and math achievement. Two studies found a positive association between larger class size and math achievement, using data from 34 OECD countries (Fung et al., 2018) and seven Confucian regions (Tan & Hew, 2019). Two studies, however, found the opposite, namely smaller class size associated with higher achievement in 48 countries (Erdogdu, 2022), and larger class size marginally negatively associated with math in the US (Pivovarova & Powers, 2019b). Contradictory results between studies, even for the same country and using the same PISA cycle, (e.g., between Pivovarova

and Powers (2019b) and Kim (2018) in the US), can be explained by the inclusion of differing predictor variables, or different analytical methods (Denny & Oppedisano, 2013).

Our review found eight relevant studies reporting mixed results regarding the relationship between student–teacher ratio and math achievement. Three studies found no association between student–teacher ratio and math achievement (e.g., Bokhove & Hampden-Thompson, 2022); two studies found mixed results that varied by students' immigrant background (Shapira, 2012) or country (Zhao & Ding, 2019); two studies found a negative association (e.g., Gamazo & Martínez-Abad, 2020), i.e., higher number of students per teacher was associated with lower math scores; and one study (Erdogdu, 2022) found a positive association (i.e., higher number of students per teacher was associated with high math scores). The study by Erdogdu (2022) was a cross-national study using pooled data from 48 countries in five PISA cycles. Caution should be taken as it was suggested that “[s]tudent-teacher ratio in sample countries needs to be reviewed” (Erdogdu, 2022, p. 18). Contradictory results between studies were even found for the same country (the US) and PISA cycle, ranging from negative (Pivovarova & Powers, 2019a) to insignificant (Kim, 2018; Zhao & Ding, 2019). Differing analytical models and approaches may explain some of these contradictory results. The nature of secondary classrooms might have made it challenging for students to identify the exact number of students per teacher. Cross-national contextual differences may also explain some contradictory findings, as student–teacher ratios tend to be larger in the high-performing East Asian countries than in other parts of the world. Similarly, low-performing students may be separated into smaller classrooms in some countries as a way to remediate their learning outcomes.

Learning time during regular school hours tended to be positively associated with math achievement in six relevant studies. Five studies found a positive association between learning time during regular school hours and math achievement (e.g., Marks & Pokropek, 2019; Santibañez & Fagioli, 2016). Shapira (2012), however, found mixed results, with learning time being positively associated with math achievement for first generation of immigrants, but insignificant for second generation immigrants and native students after family-background characteristics were controlled for. No negative association between learning time during regular school hours and math performance was found.

Teacher characteristics Teacher characteristics contained two major sub-factors, namely teachers' affective qualities (commonly measured by teacher-student relations) and teachers' qualifications/training/expertise/age. Our review found 10 relevant studies reporting mixed results regarding the relationship between teacher-student relations and math achievement. Three studies found a positive association between teachers' affective qualities and math achievement, using data from a limited number of countries (e.g., Lee, 2021). One study found a negative association, but the association was weak (Lee, 2016). Two studies found no direct association between teachers' affective qualities and math, in the US (Kitsantas et al., 2020) and in India (Areepattamannil, 2014). Four studies found mixed results depending on countries (e.g., Caro et al., 2016; Mikk et al., 2016), or depending on countries at different levels of analysis (Mikk et al., 2016). In detail, the association between teacher-student relations and math was found to be positive at the

student and school level in most countries, but it was negatively correlated at the country level (Mikk et al., 2016).

Teachers' qualifications tended to be positively associated with math achievement in seven relevant studies. Six studies found a positive association between higher level of teachers' qualifications and math (e.g., Chiu, 2015), associated with teachers having a bachelor's degree or higher (e.g., Luschei & Jeong, 2021; Orón Semper et al., 2021), a major in math (Carnoy et al., 2016; Kim, 2018), or having qualifications higher than required (Cordero & Gil-Izquierdo, 2018). One study (Liu et al., 2022) found no association, but this may be because teacher qualification was operationalized by combining formal education with years of work experience, which other studies (Carnoy et al., 2016; Cordero & Gil-Izquierdo, 2018) found it was not associated with math achievement.

Two studies found mixed results related to the association of teachers' professional development with math achievement. While Orón Semper et al. (2021) found a positive association between professional development and math using data from 79 countries, Zhao and Ding (2019) found it negative in the US and insignificant in China.

Pedagogy and assessment Pedagogy and assessment contained three major sub-factors, namely student-centered instruction, teacher-directed instruction, and cognitive activation. All 10 studies found student-centered instruction overall negatively associated with math achievement, using data from single countries (e.g., Cordero & Gil-Izquierdo, 2018) and large cross-national samples (e.g., Caro et al., 2016). However, one large cross-national study (Hermann & Kopasz, 2021) also examined sub-samples and found that the more prevalent student-centered instruction was, the higher girls' math score was.

By contrast, five relevant studies reported mixed results regarding the relationship between teacher-directed instruction and math achievement. Two cross-national studies (Caro et al., 2016; Razer et al., 2018) found a positive association in some countries and a negative association in other countries, while another cross-national studies (H. Wang et al., 2022) found a positive association in both Taiwan and Australia. Using the same PISA cycle and national context, Cordero and Gil-Izquierdo (2018) found a positive association between teacher-directed instruction and math achievement in Spain using the school as the unit of analysis, whereas Tourón et al. (2019) found no association using the student as the unit of analysis.

Our review found eight relevant studies reporting mixed results regarding the relationship between cognitive activation and math achievement. Three studies found a positive association in the US (Kitsantas et al., 2020), in Spain (Tourón et al., 2019) and in seven Confucian regions (Lee, 2021). One study found that cognitive activation was not directly associated with math but its impact through math self-efficacy and anxiety on math was significantly positive in Turkey (Yıldırım, 2012). Four studies found mixed results, depending on the countries (Caro et al., 2016; Liu et al., 2022), student race (Razer et al., 2018), country economic level (Santibañez & Fagioli, 2016), or frequency (Caro et al., 2016) or type of cognitive activation activities (Tourón et al., 2019).

Content coverage Content coverage tended to be positively associated with math achievement in 11 relevant studies, with OTL being the common terminology used across studies. Nine studies found a positive association between OTL and math achievement

(e.g., Barnard-Brak et al., 2018). Two studies found mixed results. One of them found overall OTL was positively associated with math in Russia but the sub-factors of OTL had mixed results, with exposure to formal mathematics and exposure to word problems positively associated with math but exposure to applied math was negatively associated (Carnoy et al., 2016). The other study used data from 62 countries and found OTL was positively associated with math in most countries using multiple analytical strategies, although mixed results were found in a few countries, such as Sweden (Schmidt et al., 2015). Moreover, OTL was found to interact positively with other factors impacting math outcome, such as self-efficacy (F. Wang, Wang et al., 2022) and problem-solving performance (Guo & Liao, 2022).

School characteristics

Seven major factors emerged across classes, namely school composition; school types; school location; school resources and staffing; school size; average school ICT attitude and student ICT use at school; and school prevalence of students' misbehavior. Other school factors and corresponding definitions can be found in Additional file 3: Appendix S3. School repeating/dropout rate, shortages of teachers and general staff and school prevalence of student misbehavior were consistently negatively associated with math achievement. School SES composition, general academic schools and educational resources tended to show a positive association. Results related to all the other major factors, such as school size and school location, were mixed within and across studies.

School composition School composition contained four major sub-factors, namely school SES composition; school ethnic composition; class and school gender composition; and school repeating/dropout rate.

Our review found school SES composition tended to be positively associated with math achievement in 32 relevant studies. Thirty studies found that school SES composition was positively associated with math, using large cross-national samples and multiple PISA cycles (e.g., Karakolidis et al., 2016; Sortkær & Reimer, 2018). In only two countries was the effect of school SES composition insignificant: in Turkey, after controlling for school type (Özdemir, 2016), and in Singapore (Thien & Ong, 2015), likely because school SES and school type were highly correlated in these countries. Some factors were found to weaken the association of school SES with math achievement, such as school truancy policies (Dronkers et al., 2017) and stronger instructional leadership (Gümüş et al., 2022). The strength of the association decreased as the level of school SES increased in Italy (Spagnolo et al., 2020); whether this non-linear relationship occurs in other contexts is unknown.

Our review found seven relevant studies reporting mixed results regarding the relationship between school ethnic composition and math achievement. The relationship was negative in three studies (Bokhove & Hampden-Thompson, 2022; Karakus et al., 2022; Murillo & Belavi, 2021), positive in two studies (Gamazo & Martínez-Abad, 2020; Pivovarova & Powers, 2019a), and insignificant in one study (Spagnolo et al., 2020). Zhu and Kaiser (2020) found the relationship varied across the three East Asian countries.

Our review found seven relevant studies reporting mixed results regarding the relationship between class/school gender composition and math achievement. Two studies found an insignificant association between gender composition and math in China B-S-J-G (Beijing-Shanghai-Jiangsu-Guangdong; Zhu et al., 2018) and Malaysia and Singapore (Thien & Ong, 2015). However, two studies found the impact of a higher number of girls at schools was negatively associated with math in Italy using two different PISA cycles (Meggiolaro, 2018; Spagnolo et al., 2020). One study found that the higher number of girls at schools was positively associated with math achievement using PISA 2018 data from 79 countries (Orón Semper et al., 2021). Moreover, two studies found mixed results in large cross-national studies using multiple PISA cycles, with a higher number of girls at schools positively associated with higher male students' scores, but an insignificant association with female math outcome (Munir & Winter-Ebmer, 2018; Tao & Michalopoulos, 2018).

All five studies found a negative association between school repeating/dropout rate and math achievement, in the US (Pivovarova & Powers, 2019b), Italy (Spagnolo et al., 2020), and in three cross-national studies (e.g., Orón Semper et al., 2021).

School types Our review found 18 relevant studies reporting mixed results regarding the relationship between public/private schools and math achievement. Five studies found students in public school performed better in math in the US (Pivovarova & Powers, 2019a, 2019b), Ireland (Pfeffermann & Landsman, 2011), Italy (Ferraro, 2018), and in one large cross-national sample using PISA 2012 data (Munir & Winter-Ebmer, 2018). Four studies found students in private schools performed better in math in Spain (Rodríguez et al., 2020; Tourón et al., 2019), in Australia (Dockery et al., 2020), and in one large cross-national sample using PISA 2018 data (Orón Semper et al., 2021). Three studies found insignificant relationship between private/public schools and math scores in Mexico (Aguayo-Téllez & Martínez-Rodríguez, 2020), in Spain using aggregated PISA 2009, 2012 and 2015 data (Fernández-Gutiérrez et al., 2020), in Denmark, Sweden, Finland and Iceland (e.g., Sortkær & Reimer, 2018). The remaining six studies found mixed results within their studies, across different PISA cycles (Ryan, 2013; Sousa et al., 2012), or across different analytical models (Brow, 2019) or across different countries (Kameshwara et al., 2020; Tsai et al., 2017).

It should be noted that nine studies included various measures of average school SES (e.g., Fernández-Gutiérrez et al., 2020) while the other nine did not (e.g., Aguayo-Téllez & Martínez-Rodríguez, 2020). However, the results from these two groups did not vary, meaning all the positive, negative, and insignificant relationships existed between public/private schools and math achievement, no matter if average school SES was controlled for or not. Overall, the association of public and private schools with math achievement is "hardly interpretable in a single sense" (Gamazo & Martínez-Abad, 2020, p. 9).

Six out of eight studies investigated the association of general academic schools versus vocational schools with math achievement, using data from single countries or a small number of countries. Academic schools were found to be associated with higher math scores compared to vocational schools in Italy (Meggiolaro, 2018; Spagnolo et al., 2020), Germany (Sälzer & Heine, 2016), Turkey (Özdemir, 2016), China B-S-J-G (Zhu et al., 2018), and Taiwan, Japan, South Korea, Germany and the Czech Republic (Tsai et al.,

2017). Tsai et al. (2017) did not find a relationship in the US, which is not surprising as vocational schools are rare and many secondary schools offer both academic and vocational streams.

School location Our review found 19 relevant studies reporting mixed results regarding the relationship between school location and math achievement. Eight studies found that schools located in urban areas or larger cities were positively associated with higher math scores. A positive association was found in six studies of individual countries, including Australia (Dockery et al., 2020), Spain (Fernández-Gutiérrez et al., 2020) and the US (Pivovarova & Powers, 2019a, 2019b). Two large cross-national studies (Luschei & Jeong, 2021; Orón Semper et al., 2021) also found a positive association. However, two large cross-national studies (Bokhove & Hampden-Thompson, 2022; Gümüş et al., 2022) found a negative association and one study (Bhutoria & Aljabri, 2022) found an insignificant association. Four cross-national studies found both positive and negative associations, depending on the country (e.g., Sortkær & Reimer, 2018; Zhu et al., 2018). The cross-national studies differed in their selection of PISA cycles, variables, countries and analytical approaches, which may explain why their findings differ from each other. Two studies found that the impact of school location varied according to the different analysis methods (Zhu et al., 2018), or if the factor of school type was included (Özdemir, 2016).

School resources and staffing Seventeen studies investigated school resources and staffing, which contained three major sub-factors, namely educational resources; ICT infrastructure; and shortages of teachers and general staff. Our review found educational resources tended to be positively associated with math achievement in nine relevant studies. Eight studies found a positive association (e.g., Bhutoria & Aljabri, 2022; Luschei & Jeong, 2021); one study found it insignificant (Sousa et al., 2012).

Our review found seven relevant studies reporting mixed results regarding the relationship between ICT infrastructure and math achievement. Two cross-national studies found a positive impact of ICT infrastructure (Orón Semper et al., 2021; Shapira, 2012); one study found a negative association between more instructional computers per student and math achievement in 10 OECD countries (Sousa et al., 2012); and one study found the impact of an ICT shortage in seven Confucian countries to be insignificant (Tan & Hew, 2019). Moreover, three studies (Chiu, 2020; Eickelmann et al., 2017; Hu, Gong et al., 2018) found mixed results within their respective studies. For example, aggregated school mean student ICT availability was found to be positively associated with math while other measures (such as student ICT availability at school, ratio of computers to school size) was found to be insignificant (Hu, Gong et al., 2018). School ICT availability was found to be insignificant on math after controlling for other factors, such as ICT use patterns and family SES in Taiwan (Chiu, 2020). More computer availability was found to be negative in Germany while Internet connectivity was found to be positive in Australia (Eickelmann et al., 2017).

All seven studies found a negative association between shortages of teachers/general staff and math achievement, in studies that used data from five PISA cycles in a small number of countries (e.g., Daniele, 2021; Tan & Hew, 2019) to large cross-national

studies (e.g., Bokhove & Hampden-Thompson, 2022; Luschei & Jeong, 2021). This is a very consistent finding, one of the strongest in our review.

School size Our review found 12 relevant studies reporting mixed results regarding the relationship between school size and math achievement. Seven studies found that a larger school size was significantly positively associated with math (e.g., Gümüş et al., 2022; Luschei & Jeong, 2021), while three studies found a negative association (Erdogdu, 2022; Gamazo & Martínez-Abad, 2020; Ryan, 2013). One study of three East Asian countries (Perera & Asadullah, 2019) found a positive association in Malaysia and Singapore but an insignificant impact in Korea. Sousa et al. (2012) found the association was negative using PISA 2009 data but insignificant using PISA 2006 data.

Average school ICT attitude and student ICT use at school Our review found eight relevant studies reporting mixed results regarding the relationship between ICT attitude/student use at school and math achievement. Three studies found a positive association between math achievement and use of ICT at school (Erdogdu, 2022; Ferraro, 2018), and school offering a specific program to educate students in responsible internet behavior (Orón Semper et al., 2021). Two studies, however, found a negative association between math achievement and use of ICT at school (Skryabin et al., 2015), and use of ICT in mathematics lessons involving student participation in seven Confucian regions (Tan & Hew, 2019). One study found that use of ICT at school was not associated with math achievement in Spain (Fernández-Gutiérrez et al., 2020). Two cross-national studies found mixed results that varied by country: Eickelmann et al.'s (2017) analysis of school strategies promoting ICT use and students ICT use in mathematics lessons, and Hu, Gong et al.'s (2018) analysis of school mean ICT attitudes and student ICT use at school.

School prevalence of student misbehavior The association between school prevalence of student misbehavior and math achievement was consistently negative across studies. Utilizing five PISA cycles and consisting of both single country and cross-national samples, all six studies found school prevalence of student misbehavior negatively associated with math along a range of dimensions, including prevalence of truancy (Dronkers et al., 2017; Spagnolo et al., 2020), prevalence of school bullying experience (Karakus et al., 2022), frequency of students' disruptions (Spörlein & Schlueter, 2018), and student misbehavior in general (Pivovarova & Powers, 2019b; Sousa et al., 2012).

Level 4: education systems

Three major factors emerged at Level 4, namely financial resources for education; decentralization; and tracking. Results related to all three major factors were mixed within and across studies.

Financial resources for education

Our review found nine relevant studies reporting mixed results regarding the relationship between financial resources for education and math achievement. Four

studies found a positive association (e.g., Gamazo & Martínez-Abad, 2020; Shapira, 2012). Two studies found that government funding in schools in some Asian countries (Perera & Asadullah, 2019) or regional expenditure per student in Spain (Salas-Velasco et al., 2021) was insignificantly or negatively associated with math achievement. Three studies found threshold effects, with positive associations for poor and middle-income countries but diminishing or vanished returns for higher-spending countries (Breton, 2021; Santibañez & Fagioli, 2016; Vegas & Coffin, 2015).

Decentralization

Decentralization included many sub-factors and mixed results were reported within and across 10 studies. For instance, two large cross-national studies (Breton, 2021; Han, 2018) found that school autonomy to hire teachers had an insignificant association with math achievement, but personnel autonomy was found to be positive in Singapore, negative in Malaysia and insignificant in Korea (Perera & Asadullah, 2019). These findings suggest that the association of school autonomy to hire teachers on math is likely to vary substantially across countries. Another sub-factor was school autonomy over curricula, which had a negative or insignificant association in Italy, Spain, Australia, Belgium and Canada (Daniele, 2021), or in Malaysia, Korea and Singapore (Perera & Asadullah, 2019), but a positive impact on native students in 18 OECD countries (although not on first generation of immigrants in many OECD countries) (Shapira, 2012). Moreover, autonomy for resource allocation was insignificantly associated with math achievement (Fernández-Gutiérrez et al., 2020; Perera & Asadullah, 2019), but it was positively associated with math in Costa Rica (Gimenez et al., 2018) and also on native students in OECD countries (although not on first generation of immigrants in many OECD countries) (Shapira, 2012). Increased teacher responsibility, in staffing and curricula (Luschei & Jeong, 2021) or measured by overall responsibility index (Tan, 2018), was also found to be positive. Kameshwara et al. (2020), however, found no association. Increased decision-making power among principals and government was negative (Kameshwara et al., 2020; Luschei & Jeong, 2021) but increased school board autonomy over school budgeting was positive (Luschei & Jeong, 2021).

Tracking

Our review found eight relevant studies reporting mixed results regarding the relationship between tracking and math achievement. Five studies found that tracking had a negative association with math achievement (e.g., Azzolini et al., 2012; Borgna, 2016). However, two studies found that tracking was insignificantly associated with boys' performance (Hermann & Kopasz, 2021; Scheeren, 2022). Moreover, one study found mixed results, suggesting that the impact of tracking depended on the intensity of tracking, with some ability tracking positively associated with math achievement while tracking in all subjects having detrimental effects (Cobb-Clark et al., 2012). Looking at mixed results, two studies (Hermann & Kopasz, 2021; Spörlein & Schlueter, 2018) used the same PISA 2012 math outcome in almost the same

countries (over 60 countries) but their results varied probably because the factors included in these two studies, their measure of tracking, and its data sources differed.

Level 5: macro society factors

Three major factors emerged at Level 5, namely socioeconomic country factor; national culture; and immigration and related policies. Socioeconomic country factor also contained two major sub-factors: gender equality and economic development. Results related to all major factors were mixed within and across studies.

Socioeconomic country factor

Socioeconomic country factor contained the overall concept of country SES, as well as the in-depth investigation of social development and economic development. Two out of 30 studies investigated SES using the overall country SES and both (Schmidt et al., 2015; Shapira, 2012) found overall country SES was positively associated with math achievement.

Social development—gender equality One major factor—gender equality—emerged from our review; all other factors of social development are listed in Additional file 3: Appendix S3. Our review found 11 relevant studies reporting mixed results regarding the relationship between gender equality and math performance. Six large cross-national studies found mixed results that varied by level of national level of economic development (Anghel et al., 2020), as well as PISA cycle (Eriksson et al., 2020), measure of gender equality (Reilly, 2012), or both (Stoet & Geary, 2013, 2015; Tao & Michalopoulos, 2018). On the other hand, four studies found a positive association, utilizing a range of gender equality indicators such as the Gender Gap Index (GGI; Munir & Winter-Ebmer, 2018), tertiary education and labor market (Gevrek et al., 2020), gender wage gap (Yamamura, 2019), and secondary school enrolment rates in students' ancestral country (Dockery et al., 2020). Santibañez and Fagioli (2016) found a negative association between female employment rate and math achievement, using PISA 2012 data in 50 countries, although the effect size was very small.

Economic development Our review found 16 relevant studies reporting mixed results regarding the relationship between economic development and math achievement. Eight studies found that higher economic development was significantly positively associated with math achievement in large cross-national studies using multiple PISA cycles (e.g., Chiu, 2015; Skryabin et al., 2015). However, seven studies found no association, with some PISA cycles and countries investigated overlapping with the first group of studies (e.g., Breton, 2021; Hu, Leung, & Teng, 2018). Interestingly, Daniele (2021) found that GDP per capita did not have a significant association but regional poverty rates strongly and negatively associated with math achievement. Yamamura (2019, p. 880) also found that “the positive effect of GDP decreases as the level of GDP increases”. No negative association was found among the 16 studies.

National culture

Several sub-factors of national culture were investigated among seven studies. Two sub-factors were associated with math achievement, namely Confucian culture (Breton, 2021; Tan, 2017), and long-term orientation (defined as societies which are willing to foster virtues, especially perseverance and thrift, to achieve future rewards), as found in three large cross-national studies using three PISA cycles (Breton, 2021; Fang et al., 2013; Hu, Leung, & Teng, 2018). However, other sub-factors of national culture in these three studies were found to be contradictory (individualism, uncertainty avoidance) or insignificant (masculinity).

Immigration and related policies

Our review found six relevant studies reporting mixed results regarding the relationship between immigration/related policies and math achievement. Three studies found that high immigrant concentration and selective immigration policies were positively associated with higher math achievement (Cobb-Clark et al., 2012; Shapira, 2012; Tumen, 2021). But two studies found inclusive national policies did not have a substantive impact (Arikan et al., 2017; Karakus et al., 2022). Countries with a recent history of immigration (as in many European countries) together with other educational system factors were also found to be associated with math disadvantage (Borgna, 2016). Important to point out, not all the countries were included in these six studies, with the maximum number of countries investigated was 34 OECD countries (Cobb-Clark et al., 2012). As immigrant profiles and historical trajectories vary across countries, it is likely that the association between immigration and math achievement varies substantially cross-nationally. We hypothesize that there are clear patterns that explain these cross-national differences, with immigrant status having a negative association with math achievement in Europe but a positive or insignificant association in immigrant countries such as Australia, Canada, or the US.

Discussion

The purpose of this study was to provide a comprehensive assessment of the state of knowledge about the factors driving high (and low) PISA math achievement. We followed the PRISMA 2020 guideline and conducted a systematic review on studies published since September 2011. This study found 135 factors potentially predicting PISA math achievement across studies, from specific factors (e.g., age, class size) to broader factors (e.g., family SES, national culture). About 60% of the factors were only investigated in less than five studies, mainly one or two studies. Fifty-seven factors were investigated in five or more studies and over 60% of them yielded mixed results regarding their association with math achievement. Fourteen factors were found to be positive or tend to be positive, namely student grade level, age and math self-efficacy from individual student level; overall family SES, overall home possessions, home educational resources, available books at home, and parental educational background from household context level; school SES composition, general academic schools, school educational resources, teachers' qualification, learning time during regular school hours, and content coverage from school community level. Seven factors were negative or tended

to be negative, namely student absenteeism/lack of punctuality, grade repetition, and math anxiety from individual student level; school repeating/dropout rate, school prevalence of students' misbehavior, shortage of teachers/general staff, and student-centered instruction from school community level.

Although explanatory power of factors in predicting math achievement at different levels varied across different studies (e.g., Fung et al., 2018; Martin et al., 2012), it is clear that student achievement is not attributed solely to factors at individual level, but also factors at various levels from students to societies. Four areas arising during this review are worth of discussing in depth, namely student-centered instruction, interactions among factors, possible explanations for mixed results, and suggestions for policy.

Student-centered instruction

With the growth of more student-centered instruction in many countries, it is surprising that student-centered instruction was recurrently found to be negatively associated with math achievement. It, however, echoes the results by OECD (2016, p. 71) that "...in no education system do students who reported that they are frequently exposed to enquiry-based instruction score higher in science". Several explanations are possible. First, mathematics, one of the more technical subject domains, might need more guidance from teachers to make sure students have grasped the basic understanding of mathematical concepts. Sousa et al. (2012, p. 464) even suggested a "re-thinking of whether student-centered instruction is a good approach for improving achievement, at least when it comes to the more technical topics of math and science". Second, students who participate in PISA are 15–16 years old, a majority of whom are still receiving compulsory education. Many of them may have lost interest in mathematics. For such students, a student-centered classroom without teachers' direction might not be an effective way of learning.

Third, a student-centered classroom with minimum teacher guidance might be more effective for students with sufficient foundation and capacity to manage their own learning or projects, whereas students with knowledge and skill gaps may need more guidance from teachers. Fourth, the misalignment between taught curricula and assessments might also contribute to the negative association between student-centered instruction and academic performance. Lastly, measures of student-centered instruction need to be unified. There are many components related to approaches to teaching. The balance between teachers' guidance and freedom of students' choices varies across studies depending on whether certain components are categorized to student-centered instruction or teacher-directed instruction in a given study. Measures of composite index and individual components representing student-centered instruction may also produce different results (Oliver et al., 2019). Comparisons of the findings of studies that use the same measure of student-centered instruction would be of great value.

Interactions among factors

Dividing factors into five levels provides a clear picture of each factor and its association with math achievement. However, it is very likely that the elements are intertwined and interacting with each other. The interaction among captured factors in this review occurred at different levels, with some factors interacting with each other at the same

level. For example, Han (2018) found principals from lower school SES tended to report a shortage of qualified teachers and Breton (2021) argued that without the existence of a Confucian tradition in societies, the effect of adopting a central exit exam was limited. Interaction also occurred among factors across levels. Math achievement of students in high and medium self-efficacy groups, for example, increased when class disciplinary climate improved (Cheema & Kitsantas, 2014). Due to the vast number of factors predicting math achievement, the interaction among factors has significant variation. Researchers have started to include and test the interactions among factors, such as by building path diagrams linking selected factors and their interactions with math achievement (e.g., Gabriel et al., 2020; Liu et al., 2022). However, the complex interactions between the different factors require further study.

Possible explanations for mixed results

Over 60% of major factors yielded mixed results regarding their association with math achievement within and across studies. Several reasons may explain these mixed findings. Apart from the complex interactions across factors as discussed above, they may be the result of methodological differences between studies that use different PISA cycles, selected countries, sample sizes, measures of factors, or analytical methods. One example is Sousa et al. (2012), who found that the number of cars in a family, as the indicator of household wealth, was positively associated with math achievement using PISA 2006 data but negative using PISA 2009 data. As mentioned in the results section, many studies found cross-national variations in the association between math achievement and several factors, such as gender (e.g., Kim & Law, 2012; Zhang et al., 2022), self-concept (e.g., Cheung, 2017; Thien et al., 2015), teachers' affective qualities (e.g., Caro et al., 2016; Mikk et al., 2016). This echoes the assertion of Gamazo and Martínez-Abad (2020, p. 2) that contradictory results occur "depending on the country and the PISA wave analyzed".

Regarding measures, some factors used a range of measures to test their association with math achievement. For example, among eight studies which investigated the association of early years learning experience with math achievement, three different measures of early years learning experience were utilized, namely the number of years in preschool (e.g., Cheung, 2017); whether students attended preschool (e.g., Gümüş et al., 2022); and starting first grade before six years of age (Aguayo-Téllez & Martínez-Rodríguez, 2020). Moreover, it is also unclear if PISA scales can be compared across countries (Caro et al., 2016). This agrees with the evidence provide by multiple studies (e.g., Strello et al., 2021).

Regarding analytical methods, different procedures might produce different results even when applied to the same sample. An example is Denny and Oppedisano (2013), who found a positive association between larger class size and math achievement in the US using standard ordinary least-square estimation, but an insignificant relationship when using an instrumental variables technique. Analytical procedures applied varied across studies. Threshold of confidence interval and strategies to handle missing data also varied across studies. The threshold of confidence interval to determine the statistical significance varied from $p < 0.001$ to $p < 0.10$. Reviewed studies also used various strategies to handle missing data, which may introduce bias in estimates and consequently produce distortions in the results. Overall, these agree with the claim made by

Heine and Robitzsch (2022, p. 22) that “different analytical decisions...can have a decisive influence on...cross-sectional country comparisons... in large-scale assessments...”

Mixed results may also be explained by underlying patterns related to national-cultural context or other meso- or macro-level variables. For instance, external examinations are found to be positively associated with math achievement in some national contexts but negatively associated in other contexts (Cobb-Clark et al., 2012). This might be explained by the differences in national culture across countries, since a national culture is associated with the allocation of financial resources for schools and the impact of a central exit exam (Breton, 2021). Breton (2021, p. 2) even claims that national culture is “the fundamental cause of family and school characteristics”. One of the possible reasons to explain the ambivalent results related to gender equality is that some other factors might play a more significant role in determining math achievement, which echoes the conclusion drawn by Anghel et al. (2020) that the effect of gender equality vanished when country fixed effects are accounted for. It is essential to exercise caution when interpreting these results, and it is crucial for researchers and stakeholders to contextualize the results of studies for their specific context.

Suggestions for policy

Several factors can be considered for educational policy intervention. We discuss here the factors that are amenable to policy interventions and that were (mostly)/consistently linked with math performance in our study, namely school SES composition, education resources, shortage of teachers and school staff, teachers’ qualification, and content coverage.

School SES composition was found to be mostly positively associated with math achievement. In other words, the higher the socioeconomic composition of a school, the higher the math achievement of students, regardless of their individual SES. Most policy makers and researchers in the field (e.g., Benito et al., 2014; OECD, 2019; Sciffer et al., 2022) therefore conclude that efforts to reduce school socioeconomic segregation will improve academic performance and educational equity. Educational policy makers can take several steps to address socio-economic segregation in schools. Mechanisms for allocating students to schools, such as lotteries and preference systems, can be used to create more socially balanced schools (e.g., Jang, 2022; Schwartz, 2011). In general, careful regulation of school choice policies is important as marketisation dynamics tend to exacerbate school socioeconomic segregation (Lubienski et al., 2022; OECD, 2019). Reducing between-school inequalities of human and material resources can mediate the negative impacts of school socioeconomic segregation on student achievement (OECD, 2019), as well as promote socially integrated schools as it disrupts the vicious cycle between school stratification and school segregation (Perry et al., 2022).

We found clear evidence of the importance of investing in human and material resources. Increasing investment in high quality material resources, such as library materials, laboratory equipment and audio-visual resources, may enhance learning engagement and improve students’ achievement (e.g., Areepattamannil, 2014). However, as we found mixed results regarding the association between ICT infrastructure and math achievement, further research is needed to provide policy makers with

insights. Regarding human resources, we uncovered a very consistent negative association between shortages of teachers/general staff and math achievement. Given the current severe qualified teachers and staff shortages across many countries and its potential flow-on impact on principal leadership and organizational functioning (Castro, 2023), strategies to attract and retain these critical resources should be prioritized.

We found a positive association with teachers' qualification and math achievement. In the light of this finding, authorities could require potential math teachers to hold at least a bachelor's degree and receive specialized training in math. It is worth noting that different countries have different requirements for math specialization for teachers, particularly in primary schools. Given the critical role of primary school teachers in building students' foundational mathematical knowledge, policies should prioritize the training and qualifications of these teachers to prevent students from falling behind before entering secondary school (Lagies, 2021).

Finally, we discuss content coverage. While content coverage tends to be positively associated with math achievement in this review, the sheer amount of content available to teachers and students can be overwhelming. Thus, curriculum authorities should prioritize key areas while also encouraging adaptation to local school contexts (Ryder, 2015). Incorporating math in transdisciplinary units can also increase students' exposure to math as well as promote opportunities for applying mathematical concepts and skills in a more authentic context (IBO, 2019; Moss et al., 2003).

Limitations and future directions

Several limitations of this study must be acknowledged. The first set of limitations is related to the procedure utilized in this systematic review. Even though the approach in retrieving, identifying, and analyzing studies strictly followed the PRISMA 2020 guideline, no claim of fully inclusivity can be made. First, as we limited our sample to articles published in the English language, we may have missed insights uncovered in studies published in other languages. Second, the articles included in this systematic review were only retrieved from searching in five electronic databases. Identification via other methods, such as registers, websites, organizations and citation searching, were not applied. This may have excluded some valuable studies related to the review question. Applying the condition of Q1 journal ranking to the search criteria also led to some potentially high-quality articles, which were not published in Q1 journals, being excluded. Hence, potential factors investigated to predict PISA math performance were also excluded. Furthermore, even though the interrater reliability agreement was high, there is the possibility that relevant studies could have been overlooked. Another set of limitations is related to the comparability of the review papers and their results. As mentioned in the possible explanations for mixed results, varied measures of factors, strategies to handle missing data and threshold of confidence interval applied make it challenging to compare the results across studies.

Another point which needs to be mentioned is that the list of 135 factors might not be exhaustive and there might be missing factors associated with math achievement that have been excluded or not been investigated. There might be significant variables that cannot be or have not been captured from the data, such as political influence and length of school day (e.g., Sousa et al., 2012). Another factor that was measured only

in three studies in our review is out-of-school tutoring, which has been posited as one reason for the high overall achievement in many East Asian countries (Kim, 2015; Perera & Asadullah, 2019). Furthermore, PISA collects data from students about their teachers and classrooms, but the dataset only contains student-level and school-level identifiers. This means that classroom-level analyses, for example those that examine the effect of teaching practices by a particular teacher in a particular classroom on student academic performance, are not possible.

Future research in this field could be enhanced by expanding the focus to the factors that have not attracted extensive attention but were found to be positive/negative in this review, such as student-centered instruction. More studies are also needed to investigate factors at education systems and macro society levels so that we can have a more thorough understanding of the impact of differences in national context. Studies using various research methodologies are also needed to investigate the interactions among factors at all levels and their impact on academic achievement, trying to understand the mechanism between micro-meso-macro-level factors. To complement the nature and limitation of secondary PISA data, future studies using longitudinal or experimental designs to validate the results generated by PISA analyses are needed to extend to the causality claims (Ma, 2021). In-depth qualitative work would also help unpack the mechanism among factors across levels. Moreover, researchers could consider conducting a systematic review in other major domains in PISA and compare the factors predicting math achievement and achievement in other domains to determine whether there is consistency in influential factors. Such research would provide a thorough understanding of factors predicting overall academic achievement. Systematic reviews focused on particular factors, especially those that we found to have a mixed association with math achievement, would also be useful for uncovering patterns that explain differences across studies and/or contexts. A systematic review of tracking would be particularly useful, as our study's finding of a mixed association with math achievement contrasts with the literature which mostly finds a negative association (e.g., Le Donne, 2014; OECD, 2005; Van de Werfhorst & Mijs, 2010). Lastly, considering changes caused by major events or experiences, such as the widespread impact of Covid-19 pandemic (e.g., Meinck et al., 2022) and China's crackdown on private tutoring since 2021, it would be worthwhile to examine academic achievement pre and post these events in future studies using data from PISA 2022 and further cycles.

Conclusion

This study provides a comprehensive assessment of the factors predicting math achievement in PISA based on a systematic review of over 150 high quality studies published since 2011. We find 21 factors situated at the student, family, and school levels that have a consistent association with math achievement. Some of these factors, such as overall family socio-economic status, parental educational background, and teacher qualifications, are not amenable to change in the short term. Other factors, by contrast, can be more easily monitored by students and their parents (e.g., math self-efficacy and home educational resources), or influenced by teachers and school leaders' choices (e.g., learning time during regular school hours and student-centered instruction). Furthermore, there are several factors that policy makers can address, including reducing school

socioeconomic segregation and the stratification of education resources, addressing teacher and general staff shortages, improving teacher qualifications, and promoting innovative pedagogical approaches.

Our findings further suggest that there is no universal magic bullet or easy solution for promoting high math achievement. Findings related to many factors are mixed or inconclusive, especially for factors related to education systems and the societal macro-level. This provides an impetus for further research. Some scholars claim that nations are “more likely to borrow from culturally and structurally similar systems” because it is significantly easier to import strategies from such systems (Mehta & Peterson, 2019, pp. 339–340). However, one could argue that ideological boundaries should not confine education, and schools often have beneficial practices to offer to others. Therefore, it is critical to understand national contextual differences and their interactions with factors at micro and meso levels. As “an emerging field currently situated at its take-off stage” (Hernández-Torrano & Courtney, 2021, p. 28), more work needs to be done with future PISA related research regarding macro-level factors, interactions among factors, and variety of methodologies to be applied, to name a few. Equally important, policy makers and the wider education community need to exercise extra caution when borrowing education practices and making improvement decisions, rather than automatically taking PISA results as policy solutions (Rutkowski et al., 2020).

Abbreviations

B-S-J-G	Beijing-Shanghai-Jiangsu-Guangdong
ERIC	The Education Resources Information Center
ESCS	Economic, social and cultural status
GGI	Gender gap index
ICT	Information and communication technology
ILSAs	International large-scale assessments
IRR	Interrater reliability
NCA	Necessary condition analysis
PISA	The programme for international student assessment
PRISMA	The preferred reporting items for systematic reviews and meta-analyses
OECD	Organisation for economic co-operation and development
OTL	Opportunity to learn
Q1	The top quartile
QCA	Qualitative comparative analysis
SES	Socio-economic status

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40536-023-00174-8>.

Additional file 1: Appendix S1. Search terms used in the study.

Additional file 2: Appendix S2. Summary of reviewed articles.

Additional file 3: Appendix S3. Summary of factors.

Acknowledgements

Not applicable.

Author contributions

This study is part of the first author XW's doctoral project. XW conducted the systematic review, with equal guidance from authors LP and AM and with extra support provided by TI. All the authors worked jointly on all phases of this project. XW wrote the manuscript and the other three authors provided feedback.

Funding

The authors have not received funding.

Availability of data and materials

This study is based on a systematic literature review. Methods employed are outlined in the paper. The peer-reviewed research papers that comprise the data are publicly available and listed in the references.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

We give our consent for the publication of "Factors predicting mathematics achievement in PISA: a systematic review" to be published in the *Large-scale Assessments in Education*.

Competing interests

The authors have no known competing interests to disclose.

Received: 4 January 2023 Accepted: 5 June 2023

Published online: 20 June 2023

References

- Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B., & Verger, A. (2017). The rise of international large-scale assessments and rationales for participation. *Compare*, 47(3), 434–452. <https://doi.org/10.1080/03057925.2017.1301399>
- Agirdag, O., & Vanlaar, G. (2018). Does more exposure to the language of instruction lead to higher academic achievement? A cross-national examination. *International Journal of Bilingualism*, 22(1), 123–137. <https://doi.org/10.1177/1367006916658711>
- Aguayo-Téllez, E., & Martínez-Rodríguez, F. E. (2020). Early school entrance and middle-run academic performance in Mexico: Evidence for 15-year-old students from the PISA test. *Large-scale Assessments in Education*. <https://doi.org/10.1186/s40536-020-00089-8>
- Ajello, A. M., Caponera, E., & Palmerio, L. (2018). Italian students' results in the PISA mathematics test: Does reading competence matter? *European Journal of Psychology of Education*, 33(3), 505–520. <https://doi.org/10.1007/s10212-018-0385-x>
- Anghel, B., Rodríguez-Planas, N., & Sanz-De-Galdeano, A. (2020). Is the math gender gap associated with gender equality? Only in low-income countries. *Economics of Education Review*. <https://doi.org/10.1016/j.econedurev.2020.102064>
- Areepattamannil, S. (2014). International note: What factors are associated with reading, mathematics, and science literacy of Indian adolescents? A Multilevel Examination. *Journal of Adolescence*, 37(4), 367–372. <https://doi.org/10.1016/j.adolescence.2014.02.007>
- Areepattamannil, S., Abdelfattah, F., Mahasneh, R. A., Khine, M. S., Welch, A. G., Melkonian, M., & Al Nuaimi, S. A. (2016). International note: Prediction of mathematics work ethic and performance from behavioral, normative, and control beliefs among Qatari adolescents. *Journal of Adolescence*, 46, 38–44. <https://doi.org/10.1016/j.adolescence.2015.10.016>
- Areepattamannil, S., Khine, M. S., Melkonian, M., Welch, A. G., Al Nuaimi, S. A., & Rashad, F. F. (2015). International note: Are Emirati parents' attitudes toward mathematics linked to their adolescent children's attitudes toward mathematics and mathematics achievement? *Journal of Adolescence*, 44, 17–20. <https://doi.org/10.1016/j.adolescence.2015.07.002>
- Arikan, S., Van De Vijver, F. J. R., & Yagmur, K. (2017). PISA mathematics and reading performance differences of mainstream European and Turkish immigrant students. *Educational Assessment, Evaluation and Accountability*, 29(3), 229–246. <https://doi.org/10.1007/s11092-017-9260-6>
- Avvisati, F. (2020). The measure of socio-economic status in PISA: A review and some suggested improvements. *Large-scale Assessments in Education*. <https://doi.org/10.1186/s40536-020-00086-x>
- Azzolini, D., Schnell, P., & Palmer, J. R. B. (2012). Educational achievement gaps between immigrant and native students in two "new" immigration countries: Italy and Spain in comparison. *Annals of the American Academy of Political and Social Science*, 643(1), 46–77. <https://doi.org/10.1177/0002716212441590>
- Baird, J.-A., Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G., & Yu, G. (2016). On the supranational spell of PISA in policy. *Educational Research*, 58(2), 121–138. <https://doi.org/10.1080/00131881.2016.1165410>
- Barnard-Brak, L., Lan, W. Y., & Yang, Z. (2018). Differences in mathematics achievement according to opportunity to learn: A 4pL item response theory examination. *Studies in Educational Evaluation*, 56, 1–7. <https://doi.org/10.1016/j.stueduc.2017.11.002>
- Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research*, 50(2), 837–865. <https://doi.org/10.1177/0049124118799372>
- Benito, R., Alegre, M. À., & González-Balletbò, I. (2014). School segregation and its effects on educational equality and efficiency in 16 OECD comprehensive school systems. *Comparative Education Review*, 58(1), 104–134. <https://doi.org/10.1086/672011>
- Bhutoria, A., & Aljabri, N. (2022). Patterns of cognitive returns to information and communication technology (ICT) use of 15-year-olds: Global evidence from a hierarchical linear modeling approach using PISA 2018. *Computers & Education*. <https://doi.org/10.1016/j.compedu.2022.104447>

- Bokhove, C., & Hampden-Thompson, G. (2022). Country and school family composition's effects on mathematics achievement. *School Effectiveness and School Improvement*, 33(2), 280–302. <https://doi.org/10.1080/09243453.2021.2012207>
- Borgna, C. (2016). Multiple paths to inequality. How institutional contexts shape the educational opportunities of second-generation immigrants in Europe. *European Societies*, 18(2), 180–199. <https://doi.org/10.1080/14616696.2015.1134801>
- Borgonovi, F., & Ferrara, A. (2020). Academic achievement and sense of belonging among non-native-speaking immigrant students: The role of linguistic distance. *Learning and Individual Differences*. <https://doi.org/10.1016/j.lindif.2020.101911>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328–352. <https://doi.org/10.1080/14780887.2020.1769238>
- Breton, T. R. (2021). The role of national culture in student acquisition of mathematics and reading skills. *Compare*. <https://doi.org/10.1080/03057925.2020.1870100>
- Bronfenbrenner, U. (1981). *The ecology of human development: Experiments by nature and design* (1st ed.). Cambridge: Harvard University Press.
- Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (Vol. 1, pp. 993–1028). Hoboken: John Wiley & Sons Inc.
- Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (Vol. 1, pp. 793–828). Hoboken: Wiley.
- Brow, M. V. (2019). Significant predictors of mathematical literacy for top-tiered countries/economies, Canada, and the United States on PISA 2012: Case for the sparse regression model. *British Journal of Educational Psychology*, 89(4), 726–749. <https://doi.org/10.1111/bjep.12254>
- Carnoy, M., Khavenson, T., Loyalka, P., Schmidt, W. H., & Zakharov, A. (2016). Revisiting the relationship between international assessment outcomes and educational production: Evidence from a longitudinal PISA-TIMSS sample. *American Educational Research Journal*, 53(4), 1054–1085. <https://doi.org/10.3102/0002831216653180>
- Caro, D. H., Lenkeit, J., & Kyriakides, L. (2016). Teaching strategies and differential effectiveness across learning contexts: Evidence from PISA 2012. *Studies in Educational Evaluation*, 49, 30–41. <https://doi.org/10.1016/j.stueduc.2016.03.005>
- Castro, A. J. (2023). Managing competing demands in a teacher shortage context: The impact of teacher shortages on principal leadership practices. *Educational Administration Quarterly*, 59(1), 218–250. <https://doi.org/10.1177/0013161x221140849>
- Cheema, J. R., & Kitsantas, A. (2014). Influences of disciplinary classroom climate on high school student self-efficacy and mathematics achievement: A look at gender and racial-ethnic differences. *International Journal of Science and Mathematics Education*, 12(5), 1261–1279. <https://doi.org/10.1007/s10763-013-9454-4>
- Chen, S.-Y., & Lin, S.-W. (2020). A cross-cultural study of mathematical achievement: From the perspectives of one's motivation and problem-solving style. *International Journal of Science and Mathematics Education*, 18(6), 1149–1167. <https://doi.org/10.1007/s10763-019-10011-6>
- Cheng, Q., Wang, J., Hao, S., & Shi, Q. (2014). Mathematics performance of immigrant students across different racial groups: An indirect examination of the Influence of culture and schooling. *Journal of International Migration and Integration*, 15(4), 589–607. <https://doi.org/10.1007/s12134-013-0300-x>
- Cheung, K.-C. (2017). The effects of resilience in learning variables on mathematical literacy performance: A study of learning characteristics of the academic resilient and advantaged low achievers in Shanghai, Singapore, Hong Kong, Taiwan and Korea. *Educational Psychology*, 37(8), 965–982. <https://doi.org/10.1080/01443410.2016.1194372>
- Chiu, M. M. (2015). Family inequality, school inequalities, and mathematics achievement in 65 countries: Microeconomic mechanisms of rent seeking and diminishing marginal returns. *Teachers College Record*, 117(1), 1–32. <https://doi.org/10.1177/016146811511700105>
- Chiu, M.-S. (2020). Linear or quadratic effects of ICT use on science and mathematics achievements moderated by SES: Conditioned ecological techno-process. *Research in Science & Technological Education*. <https://doi.org/10.1080/02635143.2020.1830270>
- Cobb-Clark, D. A., Sinning, M., & Stillman, S. (2012). Migrant youths' educational achievement: The role of institutions. *The ANNALS of the American Academy of Political and Social Science*, 643(1), 18–45. <https://doi.org/10.1177/0002716212440786>
- Cordero, J. M., & Gil-Izquierdo, M. (2018). The effect of teaching strategies on student achievement: An analysis using TALIS-PISA-link. *Journal of Policy Modeling*, 40(6), 1313–1331. <https://doi.org/10.1016/j.jpolmod.2018.04.003>
- Dang, D., Dearholt, S. L., Bissett, K., Ascenzi, J., & Whalen, M. (2021). *Johns Hopkins evidence-based practice for nurses and healthcare professionals* (4 ed.). Sigma Theta Tau International.
- Daniele, V. (2021). Socioeconomic inequality and regional disparities in educational achievement: The role of relative poverty. *Intelligence*. <https://doi.org/10.1016/j.intell.2020.101515>
- Denny, K., & Oppedisano, V. (2013). The surprising effect of larger class sizes: Evidence using two identification strategies. *Labour Economics*, 23, 57–65. <https://doi.org/10.1016/j.labeco.2013.04.004>
- Dockery, A. M., Koshy, P., & Li, I. W. (2020). Culture, migration and educational performance: A focus on gender outcomes using Australian PISA tests. *The Australian Educational Researcher*, 47(1), 39–59. <https://doi.org/10.1007/s13384-019-00321-7>
- Dronkers, J., Veerman, G.-J.M., & Pong, S.-L. (2017). Mechanisms behind the negative influence of single parenthood on school performance: Lower teaching and learning conditions? *Journal of Divorce & Remarriage*, 58(7), 471–486. <https://doi.org/10.1080/10502556.2017.1343558>

- Drummond, A., & Sauer, J. D. (2014). Video-games do not negatively impact adolescent academic performance in science, mathematics or reading. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0087943>
- Eickelmann, B., Gerick, J., & Koop, C. (2017). ICT use in mathematics lessons and the mathematics achievement of secondary school students by international comparison: Which role do school level factors play? *Education and Information Technologies*, 22(4), 1527–1551. <https://doi.org/10.1007/s10639-016-9498-5>
- Erdogdu, F. (2022). Potential predictors of student attainment: A longitudinal study at global level. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-022-11026-3>
- Eriksson, K., Björnstjerna, M., & Vartanova, I. (2020). The relation between gender egalitarian values and gender differences in academic achievement. *Frontiers in Psychology*, 11, 14. <https://doi.org/10.3389/fpsyg.2020.00236>
- Fan, X., Hambleton, R. K., & Zhang, M. (2019). Profiles of mathematics anxiety among 15-year-old students: A cross-cultural study using multi-group latent profile analysis. *Frontiers in Psychology*, 10, 9. <https://doi.org/10.3389/fpsyg.2019.01217>
- Fang, Z., Grant, L. W., Xu, X., Stronge, J. H., & Ward, T. J. (2013). An international comparison investigating the relationship between national culture and student achievement. *Educational Assessment, Evaluation and Accountability*, 25(3), 159–177. <https://doi.org/10.1007/s11092-013-9171-0>
- Fernández-Gutiérrez, M., Gimenez, G., & Calero, J. (2020). Is the use of ICT in education leading to higher student outcomes? Analysis from the Spanish Autonomous Communities. *Computers & Education*, 157, 15. <https://doi.org/10.1016/j.compedu.2020.103969>
- Ferraro, S. (2018). Is information and communication technology satisfying educational needs at school? *Computers & Education*, 122, 194–204. <https://doi.org/10.1016/j.compedu.2018.04.002>
- Fung, F., Tan, C. Y., & Chen, G. (2018). Student engagement and mathematics achievement: Unraveling main and interactive effects. *Psychology in the Schools*, 55(7), 815–831. <https://doi.org/10.1002/pits.22139>
- Gabriel, F., Buckley, S., & Barthakur, A. (2020). The impact of mathematics anxiety on self-regulated learning and mathematical literacy. *Australian Journal of Education*, 64(3), 227–242. <https://doi.org/10.1177/0004944120947881>
- Gamazo, A., & Martínez-Abad, F. (2020). An exploration of factors linked to academic performance in PISA 2018 through data mining techniques. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2020.575167>
- Gevrek, Z. E., Gevrek, D., & Neumeier, C. (2020). Explaining the gender gaps in mathematics achievement and attitudes: The role of societal gender equality. *Economics of Education Review*. <https://doi.org/10.1016/j.econedurev.2020.101978>
- Giannelli, G. C., & Rapallini, C. (2016). Immigrant student performance in Math: Does it matter where you come from? *Economics of Education Review*, 52, 291–304. <https://doi.org/10.1016/j.econedurev.2016.03.006>
- Giannelli, G. C., & Rapallini, C. (2019). Parental occupation and children's school outcomes in math. *Research in Economics*, 73(4), 293–303. <https://doi.org/10.1016/j.rie.2019.08.003>
- Gimenez, G., Martín-Oro, Á., & Sanaú, J. (2018). The effect of districts' social development on student performance. *Studies in Educational Evaluation*, 58, 80–96. <https://doi.org/10.1016/j.stueduc.2018.05.009>
- Gjical, K., & Lipnevich, A. A. (2021). Got math attitude? (In)direct effects of student mathematics attitudes on intentions, behavioral engagement, and mathematics performance in the US PISA. *Contemporary Educational Psychology*. <https://doi.org/10.1016/j.cedpsych.2021.102019>
- Guglielmi, R. S., & Brekke, N. (2018). A latent growth moderated mediation model of math achievement and post-secondary attainment: Focusing on context-invariant predictors. *Journal of Educational Psychology*, 110(5), 683–708. <https://doi.org/10.1037/edu0000238>
- Gümüş, S., Bellibaş, M. Ş., & Pietsch, M. (2022). School leadership and achievement gaps based on socioeconomic status: A search for socially just instructional leadership. *Journal of Educational Administration*, 60(4), 419–438. <https://doi.org/10.1108/JEA-11-2021-0213>
- Guo, S. W., & Liao, S. H. (2022). The role of opportunity to learn on student mathematics anxiety, problem-solving performance, and mathematics performance. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2022.829032>
- Gutiérrez-de-Rozas, B., López-Martín, E., & Carpintero Molina, E. (2022). Determinants of academic achievement: Systematic review of 25 years of meta-analyses. *Revista De Educación*, 398, 37–80. <https://doi.org/10.4438/1988-592X-RE-2022-398-552>
- Han, S. W. (2018). School-based teacher hiring and achievement inequality: A comparative perspective. *International Journal of Educational Development*, 61, 82–91. <https://doi.org/10.1016/j.ijedudev.2017.12.004>
- Hána, D., Hasman, J., & Kostelecká, Y. (2017). The educational performance of immigrant children at Czech schools. *Oxford Review of Education*, 43(1), 38–54. <https://doi.org/10.1080/03054985.2016.1235030>
- Heine, J.-H., & Robitzsch, A. (2022). Evaluating the effects of analytical decisions in large-scale assessments: Analyzing PISA mathematics 2003–2012. *Large-scale Assessments in Education*. <https://doi.org/10.1186/s40536-022-00129-5>
- Hermann, Z., & Kopasz, M. (2021). Educational policies and the gender gap in test scores: A cross-country analysis. *Research Papers in Education*, 36(4), 461–482. <https://doi.org/10.1080/02671522.2019.1678065>
- Hernández-Torrano, D., & Courtney, M. G. R. (2021). Modern international large-scale assessment in education: An integrative review and mapping of the literature. *Large-scale Assessments in Education*. <https://doi.org/10.1186/s40536-021-00109-1>
- Hillier, C., Zarifa, D., & Hango, D. (2021). Mind the gaps: Examining youth's reading, math and science skills across northern and rural Canada. *Rural Sociology*. <https://doi.org/10.1111/ruso.12401>
- Hoffmann, N. I. (2022). A 'win-win exercise'? The effect of westward migration on educational outcomes of Eastern European children. *Journal of Ethnic and Migration Studies*. <https://doi.org/10.1080/1369183x.2022.2087057>
- Hopfenbeck, T. N., & Görge, K. (2017). The politics of PISA: The media, policy and public responses in Norway and England. *European Journal of Education*, 52(2), 192–205. <https://doi.org/10.1111/ejed.12219>
- Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J.-A. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62(3), 333–353. <https://doi.org/10.1080/00313831.2016.1258726>

- Hu, X., Gong, Y., Lai, C., & Leung, F. K. S. (2018). The relationship between ICT and student literacy in mathematics, reading, and science across 44 countries: A multilevel analysis. *Computers & Education*, 125, 1–13. <https://doi.org/10.1016/j.compedu.2018.05.021>
- Hu, X., Leung, F. K. S., & Teng, Y. (2018). The influence of culture on students' mathematics achievement across 51 countries. *International Journal of Science and Mathematics Education*, 16(S1), S7–S24. <https://doi.org/10.1007/s10763-018-9899-6>
- Hwang, J., Choi, K. M., Bae, Y., & Shin, D. H. (2018). Do teachers' instructional practices moderate equity in mathematical and scientific literacy?: An investigation of the PISA 2012 and 2015. *International Journal of Science and Mathematics Education*, 16(S1), S25–S45. <https://doi.org/10.1007/s10763-018-9909-8>
- IBO. (2019). *Transdisciplinary learning through agency*. United Kingdom: Peterson House.
- Isphording, I. E., Piopiunik, M., & Rodríguez-Planas, N. (2016). Speaking in numbers: The effect of reading performance on math performance among immigrants. *Economics Letters*, 139, 52–56. <https://doi.org/10.1016/j.econlet.2015.11.031>
- Jang, H. (2022). *School segregation at multiple geographic levels: Implications for Educational policy*. Stanford University. ProQuest Dissertations Publishing. <https://purl.stanford.edu/ym459vv0025>
- Kameshwara, K. K., Sandoval-Hernandez, A., Shields, R., & Dhanda, K. R. (2020). A false promise? Decentralization in education systems across the globe. *International Journal of Educational Research*. <https://doi.org/10.1016/j.ijer.2020.101669>
- Karakolidis, A., Pitsia, V., & Emvalotis, A. (2016). Examining students' achievement in mathematics: A multilevel analysis of the Programme for International Student Assessment (PISA) 2012 data for Greece. *International Journal of Educational Research*, 79, 106–115. <https://doi.org/10.1016/j.ijer.2016.05.013>
- Karakus, M., Courtney, M., & Aydin, H. (2022). Understanding the academic achievement of the first- and second-generation immigrant students: A multi-level analysis of PISA 2018 data. *Educational Assessment, Evaluation and Accountability*. <https://doi.org/10.1007/s11092-022-09395-x>
- Kim, D. H., & Law, H. (2012). Gender gap in maths test scores in South Korea and Hong Kong: Role of family background and single-sex schooling. *International Journal of Educational Development*, 32(1), 92–103. <https://doi.org/10.1016/j.ijedudev.2011.02.009>
- Kim, H. (2015). Can the academic achievement of Korean students be portrayed as a product of 'shadow achievement'? *Asia Pacific Education Review*, 16(1), 119–135. <https://doi.org/10.1007/s12564-015-9361-1>
- Kim, S. (2018). ICT for children of immigrants: Indirect and total effects via self-efficacy on math performance. *Journal of Educational Computing Research*, 55(8), 1168–1200. <https://doi.org/10.1177/0735633117699954>
- King, R. B., Cai, Y., & Du, H. (2020). Societal-level utility value strengthens the relationship between student-level utility value and achievement: A person-culture fit perspective. *British Journal of Educational Psychology*, 91(1), 328–346. <https://doi.org/10.1111/bjep.12354>
- Kitsantas, A., Cleary, T. J., Whitehead, A., & Cheema, J. (2020). Relations among classroom context, student motivation, and mathematics literacy: A social cognitive perspective. *Metacognition and Learning*, 16(2), 255–273. <https://doi.org/10.1007/s11409-020-09249-1>
- Lagies, J. (2021). Orientation framework of primary school teachers teaching mathematics out-of-field– insights into a qualitative-reconstructive documentary method study. *European Journal of Teacher Education*, 44(5), 652–667. <https://doi.org/10.1080/02619768.2021.1915978>
- Le Donne, N. (2014). European variations in socioeconomic inequalities in students' cognitive achievement: The role of educational policies. *European Sociological Review*, 30(3), 329–343. <https://doi.org/10.1093/esr/jcu040>
- Lee, J. (2016). Attitude toward school does not predict academic achievement. *Learning and Individual Differences*, 52, 1–9. <https://doi.org/10.1016/j.lindif.2016.09.009>
- Lee, J. (2021). Teacher–student relationships and academic achievement in Confucian educational countries/systems from PISA 2012 perspectives. *Educational Psychology*, 41(6), 764–785. <https://doi.org/10.1080/01443410.2021.1919864>
- Lee, J., & Borgonovi, F. (2022). Relationships between family socioeconomic status and mathematics achievement in OECD and non-OECD countries. *Comparative Education Review*, 66(2), 199–227. <https://doi.org/10.1086/718930>
- Lee, J., & Stankov, L. (2013). Higher-order structure of noncognitive constructs and prediction of PISA 2003 mathematics achievement. *Learning and Individual Differences*, 26, 119–130. <https://doi.org/10.1016/j.lindif.2013.05.004>
- Lee, J., & Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences*, 65, 50–64. <https://doi.org/10.1016/j.lindif.2018.05.009>
- Lee, J., Zhang, Y., & Stankov, L. (2019). Predictive validity of SES measures for student achievement. *Educational Assessment*, 24(4), 305–326. <https://doi.org/10.1080/10627197.2019.1645590>
- Leung, F. K. S. (2014). What can and should we learn from international studies of mathematics achievement? *Mathematics Education Research Journal*, 26(3), 579–605. <https://doi.org/10.1007/s13394-013-0109-0>
- Lezhnina, O., & Kismihók, G. (2022). Combining statistical and machine learning methods to explore German students' attitudes towards ICT in PISA. *International Journal of Research & Method in Education*, 45(2), 180–199. <https://doi.org/10.1080/1743727x.2021.1963226>
- Li, H., Fortner, C. K., & Lei, X. (2015). Relationships between the use of test results and US students' academic performance. *School Effectiveness and School Improvement*, 26(2), 258–278. <https://doi.org/10.1080/09243453.2014.898662>
- Liu, X., Valcke, M., Yang Hansen, K., & De Neve, J. (2022). Does school-level instructional quality matter for school mathematics performance? Comparing teacher data across seven countries: Sustainability. <https://doi.org/10.3390/su14095267>
- Lopez-Agudo, L. A., González-Betancor, S. M., & Marcenaro-Gutierrez, O. D. (2021). Language at home and academic performance: The case of Spain. *Economic Analysis and Policy*, 69, 16–33. <https://doi.org/10.1016/j.eap.2020.11.003>
- Lubienski, C., Perry, L. B., Kim, J., & Canbolat, Y. (2022). Market models and segregation: Examining mechanisms of student sorting. *Comparative Education*, 58(1), 16–36. <https://doi.org/10.1080/03050068.2021.2013043>
- Luschei, T. F., & Jeong, D. W. (2021). School governance and student achievement: Cross-national evidence from the 2015 PISA. *Educational Administration Quarterly*, 57(3), 331–371. <https://doi.org/10.1177/0013161x20936346>

- Ma, V. J., & Ma, X. (2014). A comparative analysis of the relationship between learning styles and mathematics performance. *International Journal of STEM Education*, 1(3), 1–13. <https://doi.org/10.1186/2196-7822-1-3>
- Ma, Y. (2021). A cross-cultural study of student self-efficacy profiles and the associated predictors and outcomes using a multigroup latent profile analysis. *Studies in Educational Evaluation*. <https://doi.org/10.1016/j.stueduc.2021.101071>
- Marks, G. N., & Pokropek, A. (2019). Family income effects on mathematics achievement: Their relative magnitude and causal pathways. *Oxford Review of Education*, 45(6), 769–785. <https://doi.org/10.1080/03054985.2019.1620717>
- Martens, K., & Niemann, D. (2013). When do numbers count? The differential impact of the PISA rating and ranking on education policy in Germany and the US. *German Politics*, 22(3), 314–332. <https://doi.org/10.1080/09644008.2013.794455>
- Martin, A. J., Liem, G. A. D., Mok, M. M. C., & Xu, J. (2012). Problem solving and immigrant student mathematics and science achievement: Multination findings from the Programme for International Student Assessment (PISA). *Journal of Educational Psychology*, 104(4), 1054–1073. <https://doi.org/10.1037/a0029152>
- Matějů, P., & Smith, M. L. (2015). Are boys that bad? Gender gaps in measured skills, grades and aspirations in Czech elementary schools. *British Journal of Sociology of Education*, 36(6), 871–895. <https://doi.org/10.1080/01425692.2013.874278>
- Meggiolaro, S. (2018). Information and communication technologies use, gender and mathematics achievement: Evidence from Italy. *Social Psychology of Education*, 21(2), 497–516. <https://doi.org/10.1007/s11218-017-9425-7>
- Mehta, J., & Peterson, A. (2019). International learning communities: What happens when leaders seek to learn across national boundaries? *Journal of Educational Change*, 20(3), 327–350. <https://doi.org/10.1007/s10833-019-09348-0>
- Meinck, S., Fraillon, J., & Striehl, R. (2022). *The Impact of the COVID-19 pandemic on education: International evidence from the Responses to Educational Disruption Survey (REDS)*. UNESCO. <https://files.eric.ed.gov/fulltext/ED618542.pdf>
- Melkonian, M., Aarepattamannil, S., Menano, L., & Fildago, P. (2019). Examining acculturation orientations and perceived cultural distance among immigrant adolescents in Portugal: Links to performance in reading, mathematics, and science. *Social Psychology of Education*, 22(4), 969–989. <https://doi.org/10.1007/s11218-019-09506-5>
- Meng, L., Qiu, C., & Boyd-Wilson, B. (2019). Measurement invariance of the ICT engagement construct and its association with students' performance in China and Germany: Evidence from PISA 2015 data. *British Journal of Educational Technology*, 50(6), 3233–3251. <https://doi.org/10.1111/bjet.12729>
- Mikk, J., Kriips, H., Säälk, Ü., & Kalk, K. (2016). Relationships between student perception of teacher-student relations and PISA results in mathematics and science. *International Journal of Science and Mathematics Education*, 14(8), 1437–1454. <https://doi.org/10.1007/s10763-015-9669-7>
- Moss, D. M., Osborn, T. A., & Kaufman, D. (2003). Going beyond the boundaries. In D. Kaufman, D. M. Moss, & T. A. Osborn (Eds.), *Beyond the boundaries: A transdisciplinary approach to learning and teaching*. Westport: Praeger.
- Munir, F., & Winter-Ebmer, R. (2018). Decomposing international gender test score differences. *Journal for Labour Market Research*. <https://doi.org/10.1186/s12651-018-0246-8>
- Murillo, F. J., & Belavi, G. (2021). Differential impact of school segregation in the performance of native and non-native students in Spain. *Journal of New Approaches in Educational Research*, 10(1), 85–100. <https://doi.org/10.7821/naer.2021.1.559>
- Navarro-Martinez, O., & Peña-Acuña, B. (2022). Technology usage and academic performance in the PISA 2018 report. *Journal of New Approaches in Educational Research*, 11(1), 130–145. <https://doi.org/10.7821/naer.2022.1.735>
- Niehues, W., Kisbu-Sakarya, Y., & Selcuk, B. (2020). Motivation and maths achievement in Turkish students: Are they linked with socio-economic status? *Educational Psychology*, 40(8), 981–1001. <https://doi.org/10.1080/01443410.2020.1724887>
- Odell, B., Cutumisu, M., & Gierl, M. (2020). A scoping review of the relationship between students' ICT and performance in mathematics and science in the PISA data. *Social Psychology of Education*, 23(6), 1449–1481. <https://doi.org/10.1007/s11218-020-09591-x>
- Odell, B., Gierl, M., & Cutumisu, M. (2021). Testing measurement invariance of PISA 2015 mathematics, science, and ICT scales using the alignment method [Article]. *Studies in Educational Evaluation*. <https://doi.org/10.1016/j.stueduc.2020.100965>
- OECD. (2005). *School factors related to quality and equity: Results from PISA 2000*. OECD. <https://www.oecd.org/education/school/programme-for-international-student-assessment-pisa/34668095.pdf>
- OECD. (2016). *PISA 2015 results (Volume II): Policies and practices for successful schools*. Berlin: OECD Publishing. <https://doi.org/10.1787/9789264267510-en>
- OECD. (2019). *Balancing school choice and equity: An international perspective based on PISA*. Berlin: OECD. <https://doi.org/10.1787/2592c974-en>
- OECD. (2022). *Programme for International Student Assessment*. <https://www.oecd.org/pisa/>
- Oliver, M., McConney, A., & Woods-McConney, A. (2019). The efficacy of inquiry-based instruction in science: A comparative analysis of six countries using PISA 2015. *Research in Science Education*, 51(S2), 595–616. <https://doi.org/10.1007/s11165-019-09901-0>
- Orón Semper, J. V., Lizasoain, I., Abaurrea, J., González-García, C., & Ayuga-Téllez, E. (2021). What kind of school organizational decisions serve to enhance sustainable personal and social growth? *Sustainability*. <https://doi.org/10.3390/su13179995>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*. <https://doi.org/10.1186/s13643-016-0384-4>
- Özdemir, C. (2016). Equity in the Turkish education system: A multilevel analysis of social background influences on the mathematics performance of 15-year-old students. *European Educational Research Journal*, 15(2), 193–217. <https://doi.org/10.1177/1474904115627159>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. <https://doi.org/10.1136/bmj.n71>

- Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology, 34*(1), 29–48. <https://doi.org/10.1080/01443410.2013.797339>
- Perera, L. D. H., & Asadullah, M. N. (2019). Mind the gap: What explains Malaysia's underperformance in Pisa? *International Journal of Educational Development, 65*, 254–263. <https://doi.org/10.1016/j.ijedudev.2018.08.010>
- Perry, L. B., Rowe, E., & Lubienski, C. (2022). School segregation: Theoretical insights and future directions. *Comparative Education, 58*(1), 1–15. <https://doi.org/10.1080/03050068.2021.2021066>
- Pfeffermann, D., & Landsman, V. (2011). Are private schools better than public schools? Appraisal for Ireland by methods for observational studies. *The Annals of Applied Statistics, 5*(3), 1726–1751. <https://doi.org/10.1214/11-aos456>
- Pitsia, V., Biggart, A., & Karakolidis, A. (2017). The role of students' self-beliefs, motivation and attitudes in predicting mathematics achievement: A multilevel analysis of the Programme for International Student Assessment data. *Learning and Individual Differences, 55*, 163–173. <https://doi.org/10.1016/j.lindif.2017.03.014>
- Pivovarova, M., & Powers, J. M. (2019a). Does isolation from immigrant students benefit or harm third-plus generation students. *Education Policy Analysis Archives*. <https://doi.org/10.14507/epaa.27.4349>
- Pivovarova, M., & Powers, J. M. (2019b). Generational status, immigrant concentration and academic achievement: Comparing first and second-generation immigrants with third-plus generation students. *Large-Scale Assessments in Education, 7*(7), 1–18. <https://doi.org/10.1186/s40536-019-0075-4>
- Posso, A. (2016). Internet usage and educational outcomes among 15-year-old Australian students. *International Journal of Communication, 10*, 3851–3876.
- Radl, J., Salazar, L., & Cebolla-Boado, H. (2017). Does living in a fatherless household compromise educational success? A comparative study of cognitive and non-cognitive skills. *European Journal of Population, 33*(2), 217–242. <https://doi.org/10.1007/s10680-017-9414-8>
- Razer, M., Mittelberg, D., & Ayalon, S. (2018). The ability-track glass ceiling of Israeli schooling: Lessons from a comparative analysis of Israeli and Australian PISA 2012 data. *International Journal of Inclusive Education, 22*(2), 192–214. <https://doi.org/10.1080/13603116.2017.1362480>
- Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0039904>
- Rocher, T., & Hastedt, D. (2020). *International large-scale assessments in education: A brief guide*. IEA Compass: Briefs in Education No. 10. Amsterdam, The Netherlands: IEA. <https://eric.ed.gov/?id=ED608251>
- Rodríguez, S., Valle, A., Gironelli, L. M., Guerrero, E., Regueiro, B., & Estévez, I. (2020). Performance and well-being of native and immigrant students. Comparative analysis based on PISA 2018. *Journal of Adolescence, 85*, 96–105. <https://doi.org/10.1016/j.adolescence.2020.10.001>
- Rosa, E. M., & Tudge, J. (2013). Urie Bronfenbrenner's theory of human development: Its evolution from ecology to bioecology. *Journal of Family Theory & Review, 5*(4), 243–258. <https://doi.org/10.1111/jftr.12022>
- Rozgonjuk, D., Täht, K., & Vassil, K. (2021). Internet use at and outside of school in relation to low- and high-stakes mathematics test scores across 3 years. *International Journal of STEM Education*. <https://doi.org/10.1186/s40594-021-00287-y>
- Rutkowski, D., Thompson, G., & Rutkowski, L. (2020). Understanding the policy influence of international large-scale assessments in education. In H. Wagemaker (Ed.), *Reliability and validity of international large-scale assessment: Understanding IEA's comparative studies of student achievement*. Cham: Springer. <https://doi.org/10.1007/978-3-030-53081-5>
- Ryan, C. (2013). What is behind the decline in student achievement in Australia? *Economics of Education Review, 37*, 226–239. <https://doi.org/10.1016/j.econedurev.2013.08.008>
- Ryder, J. (2015). Being professional: Accountability and authority in teachers' responses to science curriculum reform. *Studies in Science Education, 51*(1), 87–120. <https://doi.org/10.1080/03057267.2014.1001629>
- Salas-Velasco, M., Moreno-Herrero, D., & Sánchez-Campillo, J. (2021). Positive geographical spillovers of human capital on student learning outcomes. *Applied Spatial Analysis and Policy, 14*(2), 415–443. <https://doi.org/10.1007/s12061-020-09366-z>
- Salkind, N. (2011). Interrater reliability. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics*. Thousand Oaks: Sage Publications. <https://doi.org/10.4135/9781412952644>
- Sälzer, C., & Heine, J.-H. (2016). Students' skipping behavior on truancy items and (school) subjects and its relation to test performance in PISA 2012. *International Journal of Educational Development, 46*, 103–113. <https://doi.org/10.1016/j.ijedudev.2015.10.009>
- Santibañez, L., & Fagioli, L. (2016). Nothing succeeds like success? Equity, student outcomes, and opportunity to learn in high- and middle-income countries. *International Journal of Behavioral Development, 40*(6), 517–525. <https://doi.org/10.1177/0165025416642050>
- Scheeren, L. (2022). The differential impact of educational tracking on SES gaps in educational achievement for boys and girls. *European Sociological Review*. <https://doi.org/10.1093/esr/jcac012>
- Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The role of schooling in perpetuating educational inequality: An international perspective. *Educational Researcher, 44*(7), 371–386. <https://doi.org/10.3102/0013189x15603982>
- Schmidt, W. H., Guo, S., & Houang, R. T. (2021). The role of opportunity to learn in ethnic inequality in mathematics. *Journal of Curriculum Studies, 53*(5), 579–600. <https://doi.org/10.1080/00220272.2020.1863475>
- Schwartz, H. (2011). Housing policy is school policy: Economically integrative housing promotes academic success in Montgomery County MD. *The Education Digest, 76*(6), 42.
- Sciffer, M. G., Perry, L. B., & McConney, A. (2022). Does school socioeconomic composition matter more in some countries than others, and if so, why? *Comparative Education, 58*(1), 37–51. <https://doi.org/10.1080/03050068.2021.2013045>
- Shapira, M. (2012). An exploration of differences in mathematics attainment among immigrant pupils in 18 OECD countries. *European Educational Research Journal, 11*(1), 68–95. <https://doi.org/10.2304/eej.2012.11.1.68>

- Skryabin, M., Zhang, J., Liu, L., & Zhang, D. (2015). How the ICT development level and usage influence student achievement in reading, mathematics, and science. *Computers & Education*, 85, 49–58. <https://doi.org/10.1016/j.compedu.2015.02.004>
- Sortkær, B., & Reimer, D. (2018). Classroom disciplinary climate of schools and gender – evidence from the Nordic countries. *School Effectiveness and School Improvement*, 29(4), 511–528. <https://doi.org/10.1080/09243453.2018.1460382>
- Sousa, S., Park, E. J., & Armor, D. J. (2012). Comparing effects of family and school factors on cross-national academic achievement using the 2009 and 2006 PISA surveys. *Journal of Comparative Policy Analysis: Research and Practice*, 14(5), 449–468. <https://doi.org/10.1080/13876988.2012.726535>
- Spagnolo, F. S., Salvati, N., D'Agostino, A., & Nicaise, I. (2020). The use of sampling weights in M-quantile random-effects regression: An application to Programme for International Student Assessment mathematics scores. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(4), 991–1012. <https://doi.org/10.1111/rssc.12418>
- Spörlein, C., & Schlueter, E. (2018). How education systems shape cross-national ethnic inequality in math competence scores: Moving beyond mean differences. *PLoS ONE*, 13(3), 21. <https://doi.org/10.1371/journal.pone.0193738>
- Stoet, G., Bailey, D. H., Moore, A. M., & Geary, D. C. (2016). Countries with higher levels of gender equality show larger national sex differences in mathematics anxiety and relatively lower parental mathematics valuation for girls. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0153857>
- Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0057988>
- Stoet, G., & Geary, D. C. (2015). Sex differences in academic achievement are not related to political, economic, or social equality. *Intelligence*, 48, 137–151. <https://doi.org/10.1016/j.intell.2014.11.006>
- Strello, A., Strietholt, R., Steinmann, I., & Siepmann, C. (2021). Early tracking and different types of inequalities in achievement: Difference-in-differences evidence from 20 years of large-scale assessments. *Educational Assessment, Evaluation and Accountability*, 33(1), 139–167. <https://doi.org/10.1007/s11092-020-09346-4>
- Sulis, I., Giambona, F., & Porcu, M. (2020). Adjusted indicators of quality and equity for monitoring the education systems over time. Insights on EU15 countries from PISA surveys. *Socio-Economic Planning Sciences*. <https://doi.org/10.1016/j.seps.2019.05.005>
- Tan, C. Y. (2015). The contribution of cultural capital to students' mathematics achievement in medium and high socioeconomic gradient economies. *British Educational Research Journal*, 41(6), 1050–1067. <https://doi.org/10.1002/BERJ.3187>
- Tan, C. Y. (2017). Do parental attitudes toward and expectations for their children's education and future jobs matter for their children's school achievement? *British Educational Research Journal*, 43(6), 1111–1130. <https://doi.org/10.1002/berj.3303>
- Tan, C. Y. (2018). Examining school leadership effects on student achievement: The role of contextual challenges and constraints. *Cambridge Journal of Education*, 48(1), 21–45. <https://doi.org/10.1080/0305764x.2016.1221885>
- Tan, C. Y., & Hew, K. F. (2019). The impact of digital divides on student mathematics achievement in Confucian heritage cultures: A critical examination using PISA 2012 data. *International Journal of Science and Mathematics Education*, 17(6), 1213–1232. <https://doi.org/10.1007/s10763-018-9917-8>
- Tao, H.-L., & Michalopoulos, C. (2018). Gender equality and the gender gap in mathematics. *Journal of Biosocial Science*, 50(2), 227–243. <https://doi.org/10.1017/s0021932017000141>
- Teig, N., Scherer, R., & Olsen, R. V. (2022). A systematic review of studies investigating science teaching and learning: Over two decades of TIMSS and PISA. *International Journal of Science Education*, 44(12), 2035–2058. <https://doi.org/10.1080/09500693.2022.2109075>
- Thien, L. M., Darmawan, I. G. N., & Ong, M. Y. (2015). Affective characteristics and mathematics performance in Indonesia, Malaysia, and Thailand: What can PISA 2012 data tell us? *Large-Scale Assessments in Education*, 3(3), 1–16. <https://doi.org/10.1186/s40536-015-0013-z>
- Thien, L. M., & Ong, M. Y. (2015). Malaysian and Singaporean students' affective characteristics and mathematics performance: Evidence from PISA 2012. *Springerplus*. <https://doi.org/10.1186/s40064-015-1358-z>
- Thompson, G., Sellar, S., & Lingard, B. (2016). The life of data: Evolving national testing. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment*. Abingdon: Routledge. <https://doi.org/10.4324/9781315659312-16>
- Tourón, J., Navarro-Asencio, E., Lizasoain, L., López-González, E., & García-San Pedro, M. J. (2019). How teachers' practices and students' attitudes towards technology affect mathematics achievement: Results and insights from PISA 2012. *Research Papers in Education*, 34(3), 263–275. <https://doi.org/10.1080/02671522.2018.1424927>
- Tsai, S.-L., Smith, M. L., & Hauser, R. M. (2017). Families, schools, and student achievement inequality: A multilevel MIMIC model approach. *Sociology of Education*, 90(1), 64–88. <https://doi.org/10.1177/0038040716683779>
- Tumen, S. (2021). The effect of refugees on native adolescents' test scores: Quasi-experimental evidence from PISA. *Journal of Development Economics*. <https://doi.org/10.1016/j.jdeveco.2021.102633>
- Ünal, E., Uzun, A. M., & Kilis, S. (2022). Does ICT involvement really matter? An investigation of Turkey's case in PISA 2018. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-022-11067-8>
- Van de Werfhorst, H. G., & Mijts, J. J. B. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, 36(1), 407–428. <https://doi.org/10.1146/annurev.soc.012809.102538>
- Vegas, E., & Coffin, C. (2015). When education expenditure matters: An empirical analysis of recent international data. *Comparative Education Review*, 59(2), 289–304. <https://doi.org/10.1086/680324>
- Wang, F., Liu, Y., & Leung, S. O. (2022). Disciplinary climate, opportunity to learn, and mathematics achievement: An analysis using doubly latent multilevel structural equation modeling. *School Effectiveness and School Improvement*. <https://doi.org/10.1080/09243453.2022.2043393>
- Wang, F., Wang, Y., Liu, Y., & Leung, S. O. (2022). Investigating the measurement of OTL in PISA 2012 and its relationship with self-efficacy and mathematics achievement: Doubly latent multilevel analyses. *Scandinavian Journal of Educational Research*. <https://doi.org/10.1080/00313831.2022.2070929>

- Wang, H.-H., Hong, Z.-R., She, H.-C., Smith, T. J., Fielding, J., & Lin, H.-S. (2022). The role of structured inquiry, open inquiry, and epistemological beliefs in developing secondary students' scientific and mathematical literacies. *International Journal of STEM Education*. <https://doi.org/10.1186/s40594-022-00329-z>
- Wu, Y.-J., Carstensen, C. H., & Lee, J. (2020). A new perspective on memorization practices among East Asian students based on PISA 2012. *Educational Psychology*, 40(5), 643–662. <https://doi.org/10.1080/01443410.2019.1648766>
- Xiao, F., & Sun, L. (2021). Students' motivation and affection profiles and their relation to mathematics achievement, persistence, and behaviors. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2020.533593>
- Xiaomin, L., & Auld, E. (2020). A historical perspective on the OECD's 'humanitarian turn': PISA for Development and the learning framework 2030. *Comparative Education*, 56(4), 503–521. <https://doi.org/10.1080/03050068.2020.1781397>
- Xie, C., & Ma, Y. (2019). The mediating role of cultural capital in the relationship between socioeconomic status and student achievement in 14 economies. *British Educational Research Journal*, 45(4), 838–855. <https://doi.org/10.1002/berj.3528>
- Yamamura, E. (2019). Gender wage gap and its effect on test scores of immigrant students. *Journal of Economic Studies*, 46(4), 872–887. <https://doi.org/10.1108/jes-04-2018-0141>
- Yıldırım, S. (2012). Teacher support, motivation, learning strategy use, and achievement: A multilevel mediation model. *The Journal of Experimental Education*, 80(2), 150–172. <https://doi.org/10.1080/00220973.2011.596855>
- Zhang, Y., Russell, S., & Kelly, S. (2022). Engagement, achievement, and teacher classroom practices in mathematics: Insights from TIMSS 2011 and PISA 2012. *Studies in Educational Evaluation*. <https://doi.org/10.1016/j.stueduc.2022.101146>
- Zhao, Y., & Ding, C. (2019). The association between students' mathematics knowledge and factors related to students, parents, and school: A cross-cultural comparison study. *International Journal of Educational Research*, 93, 210–217. <https://doi.org/10.1016/j.ijer.2018.11.006>
- Zhu, Y., & Kaiser, G. (2020). Do East Asian migrant students perform equally well in mathematics? *International Journal of Science and Mathematics Education*, 18(6), 1127–1147. <https://doi.org/10.1007/s10763-019-10014-3>
- Zhu, Y., Kaiser, G., & Cai, J. (2018). Gender equity in mathematical achievement: The case of China. *Educational Studies in Mathematics*, 99(3), 245–260. <https://doi.org/10.1007/s10649-018-9846-z>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
