International Journal of Science Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tsed20

Worldwide predictors of science literacy in lower-secondary students: a TIMSS 2019 analysis

João Marôco, Heidi Harju-Lukkainnen & Juhani Rautopuro

To cite this article: João Marôco, Heidi Harju-Lukkainnen & Juhani Rautopuro (11 Sep 2024): Worldwide predictors of science literacy in lower-secondary students: a TIMSS 2019 analysis, International Journal of Science Education, DOI: 10.1080/09500693.2024.2394239

To link to this article: https://doi.org/10.1080/09500693.2024.2394239

Published online: 11 Sep 2024.

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION
https://doi.org/10.1080/09500693.2024.2394239

Worldwide predictors of science literacy in lower-secondary students: a TIMSS 2019 analysis

João Marôco [a], Heidi Harju-Lukkainnen [b,c] and Juhani Rautopuro [d]

[a]William James Centre for Research, ISPA – Instituto Universitrio, Lisbon, Portugal; [b]University of Jyväskylä, Jyväskylä, Finland; [c]Nord University, Bodø, Norway; [d]Finnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, Finland

ABSTRACT

In today's technologically advanced and environmentally challenging landscape, enhancing global scientific literacy is increasingly imperative. A solid grasp of scientific principles is vital for fostering innovation and achieving sustainable development worldwide. Utilizing data from the International Association for the Evaluation of Educational Achievement (IEA) Trends in International Mathematics and Science Study (TIMSS), this study inquiries into various factors impacting grade eight students' science literacy in the TIMSS 2019 assessment. Exploring student socioeconomic status, cultural background, attitudes towards science, school organisation, teacher experiences, and country-level variables, we aimed to identify predictive associations with science literacy. Additionally, we correlated TIMSS 2019 results with the United Nations Development Program's Human Development Index (HDI). Employing three-level Hierarchical Linear Modeling, we assessed which student, teacher, and country-level variables were linked to grade 8 science literacy. Results indicate that the students' socioeconomic resources and motivation/confidence in science positively influence science literacy. Conversely, a higher percentage of science items covered in national curricula was weakly associated with lower science literacy scores. Teachers reporting greater teaching limitations due to student needs showed negative correlations with science scores. At the country level, the HDI emerged as the strongest correlate with grade 8 science literacy. These findings underscore the multifaceted nature of science literacy and its interplay with various socio-cultural and educational factors on a global scale.

ARTICLE HISTORY
Received 11 May 2024
Accepted 15 August 2024

KEYWORDS
Science literacy; TIMSS 2019; human development index

Introduction

The significance of science education in equipping students for the modern world was highlighted by the OECD already in 1996 (OECD 1996). It identified a transformation in the working field in developing countries, highlighting the performance requirements of the future, reflecting how knowledge is created and utilised in a modern society. In all

CONTACT João Marôco jpmaroco@gmail.com William James Centre for Research, ISPA – Instituto Universitrio, Lisbon, Portugal

this, scientific literacy plays a pivotal role in today's societies by driving innovation, progress, and understanding of the world around us. From advancements in medicine that save lives to technological breakthroughs that enhance communication and improve standards of living, science permeates every aspect of contemporary life. It provides evidence-based solutions to complex problems, informing policy-making and guiding decision-making processes. Science serves as the cornerstone of progress, offering insights and solutions that shape the trajectory of societies towards a more sustainable future. This perspective on science was adopted already in 1999 at the UNESCO's 'Science and the use of Scientific Knowledge Declaration' (UNESCO 1999). From these premises, quality science education of the youngest generations is a much-required feature of modern educational systems (see, e.g. van Driel et al. 2012). Scientific literacy – defined as the understanding of scientific concepts, applying scientific knowledge, and using scientific inquiry skills to engage with science-related issues, and with the ideas of science, as a reflective citizen (see, e.g. OECD 2023) – empowers individuals to critically evaluate information, fostering a more informed and engaged citizenry. As noted by Dickman et al. (2009), Hodson (2003) and many others, science education is essential for preparing the workforce of the future and ensuring global competitiveness and connected citizenship in a constantly changing world.

There has been a significant growth of international large-scale assessments (ILSA) since the 1990 and their power for change has been observed worldwide. According to Johansson and Yang Hansen (2019), ILSA can provide a catalyst for education change, and inspire the search for models from other countries which might solve perceived educational problems. They argue further that ILSA have gained a prominent position in societal discourse and seem to play an increasingly important role in decision-making and school reforms. According to Fischman et al. (2019) ILSA promoted, in many countries, new conditions for educational comparison at the national, regional, and global levels. Further, according to Johansson (2016), there is a considerable amount of literature discussing the uses and consequences of large-scale assessments. However, it is also apparent that large-scale international assessments can be a valuable resource for studying global trends and evolving systems in education, but the problem lies in the fact that educational research all too often appears underused in the discussions.

From these premises, the International Association for the Evaluation of Educational Achievement (IEA) Trends in International Mathematics and Science Study (TIMSS) is a vital tool for understanding the educational landscape worldwide, particularly not only in the realm of science education, but also in shaping the education systems further. Conducted every four years since 1995, TIMSS assesses the knowledge and skills of fourth- and eighth-grade students in mathematics and science (IEA 2024). While grade four is a keystone in foundational knowledge, language integration, curiosity, and wonder, grade 8 is a pivotal stage where curiosity meets foundational knowledge and critical thinking, shaping young scientists who will contribute to our ever-evolving world. TIMSS provides valuable insights into the effectiveness of science education systems across countries, identifying strengths, weaknesses, and areas for improvement. By bench-marking performance against international standards, TIMSS helps policymakers, educators, and researchers make informed decisions to enhance science education curricula, teaching methods, and educational policies. It serves as a catalyst for global collaboration and knowledge exchange, fostering dialogue and sharing best practises to elevate science education standards worldwide (IEA 2024).

The TIMSS 2019 Science Grade 8 framework provides a comprehensive structure for assessing students' scientific knowledge and skills. Organised around two key dimensions, Content and Cognitive, it ensures a robust evaluation of science education across different countries and cultures (Mullis & Martin 2017). The Content Dimension encompasses four primary domains: Biology, accounting for 35% of the assessment time; Chemistry, constituting 20% of the assessment time; Physics, constituting 20% of the assessment time; and Earth Science making up 20% of the assessment time. The Cognitive domain assesses the students' cognitive abilities, namely, the students' abilities to apply scientific reasoning, analyze data, and solve problems. By emphasizing both content knowledge and cognitive skills, the TIMSS 2019 Science Framework aims to prepare students for informed decision-making, scientific literacy, and future careers in science, technology, and engineering (see Mullis & Martin 2017, for a comprehensive description of the TIMSS 2019 Science grade 8 framework). Students participating at TIMSS 2019 were selected by a robust multistage random sampling design ensuring representation across participating countries. About half of the countries administered the assessment via computer (eTIMSS), while others used the traditional paper-and-pencil format. The transition to the computer format posed additional challenges and mode effects, but TIMSS used advanced statistical linking methods to enable comparable trend reporting for both modes (Foy et al., 2020). Selected students took the test in controlled conditions under rigorous quality assurance procedures to maintain consistency and reliability. In grade 8, students took a 90-minute test (assessing the four content domains) and a 30-min questionnaire regarding their demographics context, schooling, personal experiences, motivation, and expectations (Mullis & Martin 2017). Due to the content extension and time limitation, not all students take the same test. Tests are organised in random blocks of content with planned missing responses imputed using Gaussian population models derived from IRT scaling and weighted regression on socioeconomic and demographic variables. Linking across modes (paper and computer) with previous TIMSS results were also performed (von Davier 2020). Five plausible values for the science literacy of each student were estimated as random draws from an empirically derived distribution of population score values. The estimation process takes into account the student's observed responses to assessment items and background variables. Plausible values allow for a more robust assessment by considering the uncertainty associated with each student's performance. They help address the limitation of administering a limited number of test items to each student, ensuring that individual content-related scale scores are still meaningful. Each plausible value thus represents what an individual's performance on the entire assessment might have been if it had been observed (Martin et al. 2020).

In addition to the student test and questionnaire, TIMSS also applies context questionnaires to school principals, Science and Math teachers. From their responses a set of psychometric scales and indices are derived (Fishbein et al. 2021). The scales and individual variables used in this research are described in the Methods Data section (see below). National curricula matching with the TIMSS assessed curriculum was also reported by the National Research Coordinators in each participating country.

In this study, we probed student socioeconomic and cultural status, actions and attitudes towards science, schools' organisational characteristics and teachers' experiences and motivations, as well as country variables to seek predictive associations with grade eight students' science literacy as assessed by TIMSS. Despite the extensive research on the factors influencing student achievement in science, there is a significant gap in understanding how the interplay of individual, school, and national-level variables collectively impacts science literacy. Previous studies have often focused on isolated factors, such as socioeconomic status or teacher quality, without considering the complex interactions between these variables (Ersan & Rodriguez 2020, e.g.). Moreover, there is limited comparative research that examines these relationships across but a few countries, particularly using robust international assessments like TIMSS (Kaya 2022). This study aims to fill this gap by providing a comprehensive analysis that integrates multiple dimensions of influence (students, families, teachers, schools and countries), offering a more holistic understanding of what drives science literacy among grade eight students globally. Therefore, our research hypotheses are the following:

H_1: Countries' human development, expenditure in education, TIMSS science curricula coverage by the national curriculum, and science assessment practises (e.g. exams and school evaluations) can explain the students' science literacy.
H_2: Students' socioeconomic and cultural resources, as well as the students' confidence and attitudes towards science are positively associated with the students' science literacy
H_3: Both school socioeconomic characteristics and organisational values can explain the students' science literacy.
H_4: Teachers' experience and academic training as well as teachers' attitudes towards teaching science are positively associated with the students' science literacy

Methods

Data

Science literacy for grade 8 around the world was gathered from the TIMSS 2019 edition. This assessment is carried out by the International Association for the Evaluation of Educational Achievement (IEA), an autonomous international collaboration of national research institutions and government bodies dedicated to cross-national educational evaluations. In 2019, TIMSS at grade 8 encompassed data from 38 countries, making it one of the most extensive international assessments worldwide. Apart from measuring student achievement, TIMSS also provides insights into various factors related to family and school environments. In the 2019 TIMSS edition, questionnaires were administered to students, schools, and teachers. The data on science literacy utilised in this research are openly accessible through the TIMSS and PIRLS International Study Center, which is housed at the Lynch School of Education, Boston College, Newton, MA (https://timss2019.org/international-database/). For a comprehensive understanding of the available data files and variables, one can refer to Yin and Fishbein (2020) and Fishbein et al. (2021).

Student-level variables (Level 1)

From the questionnaire to the students, indices were derived using partial credit IRT models (see Appendix 16B in Yin & Fishbein, 2020). In addition to the student's sex (BSBG01), the following scales were used in this research:

- *Home Educational Resources* (BSBGHER): a scale derived from students reports regarding the number of books (five categories), number of home study supports (internet access, own room, etc.), and parents' highest education level. Cronbach's α for this scale ranged from 0.25 for Japan up to 0.64 for Turkey with an overall average of 0.43 ($SE = 0.010$).
- *Sense of School Belonging* (BSBGSSB): a scale derived from students responses to five items on a four-point rating scale (from 'Disagree a lot' to 'Agree a lot') addressing feelings and attitudes towards school belong (e.g. I like being in school; I am proud to go to this school). Cronbach's α ranged from 0.70 in South Africa to 0.86 in Hong Kong SAR, with an overall mean of 0.80 ($SE = 0.006$).
- *Student Bullying* (BSBGSB): a scale derived from the students' responses to fourteen items that assessed the frequency, on a four-point rating scale (from never to at least once a week), of bullying behaviors and attitudes (e.g. Shared embarrassing photos of me online; Physically hurt me). Cronbach's α ranged from 0.78 for Kazakhstan up to 0.92 for Jordan and Qatar, with an overall mean of 0.87 ($SE = 0.006$).
- *Students Like Learning Science* (SBSBGSLS): a scale derived from students responses to nine items that assess the agreement, on a four-point scale, of items related to learning and enjoying science (e.g. I enjoy learning science; Science teaches me how things in the world work). Average Cronbach's α was 0.90 ($SE = 0.005$) ranging from 0.83 for Egypt to 0.93 for Australia, England, and Ireland. This scale was not applied by 12 of the 38 participating countries and regions.
- *Instructional Clarity in Science Lessons* (BSBGICS): a scale derived from students responses to seven items, on a four-point agreement scale, that reflected teachers' expectations towards students (e.g. I know what my teacher expects me to do; My teacher is easy to understand). Cronbach's α ranged from 0.83 (Egypt) to 0.95 (in Hong Kong SAR). Mean α was 0.90 ($SE = 0.006$). This scale was also not applied by 12 of the 38 participants.
- *Students Confident in Science* (BSBGSCS): a scale created based on students responses to eight items that assessed their degree of confidence (from disagree a lot to Agree a lot) in Science (e.g. I usually do well in science; I learn things quickly in science). This scale was not applied by 12 participants. For the remaining 26 participants, Cronbach's α was 0.83 ($SE = 0.012$). The Minimum value of α was 0.73 (Egypt) and the maximum was 0.92 (Korea).
- *Students Value Science* (BSBGSVS): scale derived from students responses on a four-point agreement scale to nine items expressing their valorisation of science (e.g. I think learning science will help me in my daily life; It is important to do well in science). The mean Cronbach's α for the 38 participants was 0.92 ($SE = 0.002$) with a minimum value of 0.89 (Oman) and a maximum of 0.94 (Australia).
- *Self-Efficacy for Computer Use* (BSBGSEC): a scale derived from students responses to seven items addressing familiarity and efficacy with computers (e.g. I am good at using the computer; It is easy for me to find information on the internet) on a four-point

rating scale (from 1– Disagree a lot to 4– Agree a lot). Higher scores indicate higher self-efficacy for computer use. The average Cronbach's α for the 22 countries that applied the scale was 0.816 ($SE = 0.008$) ranging from 0.76 (Italy) to 0.89 (Rep. of Korea).

Schools and Science Teacher variables (Level 2)

These are variables and scales derived from the school's principal and science teachers' questionnaires.

- *Percentage of Economic disadvantaged students* (BCBG03A): A four-point ordinal variable (1– 0 to 10%; 2– 11 to 25%; 3– 26 to 50%, 4– More than 50%) used as interval variable in the HLM analysis.
- *Instructional days per year* (BCBG06A): a ratio variable accounting for the total number of days of instruction.
- *Existing Science Laboratories* (BCBG08A): a nominal variable stating the existence of science labs in schools, recoded to 0– No, 1– Yes.
- *Existing school library* (BCBG10A): a nominal variable stating the existence of a library in schools. Recoded to 0– No, 1– Yes.
- *School Discipline Problems* (BCBGDAS): a scale derived from principals responses to eleven items (e.g. Arriving late at school; Cheating) that assessed the severity of disciplinary problems (from 1 – Severe problem to 4 – Not a problem). Higher values on the scale indicate fewer discipline problems. Mean Cronbach α was 0.91 ($SE = 0.008$) ranging from 0.81 (Hong Kong SAR) to 0.98 (Kazakhstan)
- *School Emphasis on Academic Success* (BCBGEAS): a scale derived from principals responses to eleven items describing the school's stakeholders' emphasis on students' achievement (e.g. Teachers understanding of the school's curricular goals; Parental support for students) on a 5-point rating scale (1 – Very low; 5 – Very high). Higher scale scores indicate higher levels of emphasis. There are no published reports on the reliability of this scale.
- *Instruction Affected by Science Resource Shortages* (BCBGSRS): scale derived from principals responses to thirteen items assessing the shortage of science teaching resources (e.g. Instructional materials (e.g. textbooks); Teachers with a specialisation in science) on a rating scale from 1-Not at all to 4-A lot. Higher values on the scale indicate higher affection for instruction. There are no published reports on the reliability of this scale.
- *Years been teaching (BTBG01): a ratio variable for the duration of the teaching years reported by science teachers.When more than one* teacher per school was sampled, the median of all teachers was used.
- *Level of formal education completed* (BTBG04): an ordinal variable for the level of education of teachers (from 1 – Did not complete to 7 – Doctoral or equivalent level). Treated as a metric variable in the HLM analysis. When more than one teacher per school was sampled, the median of all teachers was used.
- *Science as major area of study* (BTBG05G): a nominal variable registering if a teacher majored in science or not. Recoded to 1 – Yes, 0 – No. When more than one teacher per school was sampled, the median of all teachers was used.

- *Teaching Limited by Student Needs* (BTBGLSN): a scale created based on teachers responses to eight items that assessed students' needs and readiness (e.g. Students lacking prerequisite knowledge or skills; Uninterested students) on a rating scale of 4 – Not at all to 1 – A lot. Higher scale values indicate lower limitations due to students needs. Mean Cronbach's α was 0.78 ($SE = 0.008$) ranging from 0.68 (Egypt) to 0.86 (Israel). When more than one teacher per school was sampled, the median of all teachers was used.
- *Teachers Job Satisfaction* (BTBGTJS): a scale derived from teachers responses to five items addressing job satisfaction (e.g. I am content with my profession as a teacher; I find my work full of meaning and purpose) on a rating scale of 1 – Never or almost never to 4 – Very often. Higher values of the scale indicate higher satisfaction. Mean Cronbach's α was 0.89 ($SE = 0.006$) ranging from 0.78 (United States) to 0.95 (Singapore). When more than one teacher per school was sampled, the median of all teachers was used.
- *Teachers Emphasis in Science Investigation* (BTBSESI): a scale created based on teachers responses to eight items that assessed frequency of scientific inquiry in class (e.g. Design or plan experiments or investigations; Do field work outside the class) on a rating scale of 1 – Never to 4 – Every or almost every lesson. Higher scale values indicate higher emphasis. Mean Cronbach's α was 0.84 ($SE = 0.006$) ranging from 0.76 (England) to 0.91 (Turkey). When more than one teacher per school was sampled, the median of all teachers was used.

Country -level variables (Level 3)

At the country level, the following variables were probed:

- *Human Development Index* (HDI): The UNDP Human Development Index (HDI) is a composite measure of human well-being, combining indicators of health, education, and standard of living. It assigns a score between 0 and 1 to countries based on life expectancy, education levels, and income per capita. The HDI is available at https://hdr.undp.org/data-center/human-development-index#/indicies/HDI
- *Expenditure in Education* (ExpEdu): The Total government expenditure on education, as a percentage of the gross domestic product (GDP). Data retrieved from the World Bank at https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS
- *Coverage of TIMSS Science Curriculum by the National Curriculum* (CurrCov): Percentage of items in the Science grade 8 test that is covered by the national curriculum as reported by TIMSS countries coordinators. Data retrieved from the TIMSS Curriculum Match Analysis is available at the TIMSS and PIRLS International Study Center.
- *National or regional Exams* (Exam): a nominal variable reporting the existence of national or regional high-stakes exams (0 – No, 1 – Yes). Data reported by TIMSS countries coordinators and available at the TIMSS 2019 Encyclopedia.
- *Schools conduct Science self-evaluations* (SchSelfEv): A nominal variable reporting the existence of schools science evaluations (0 – No, 1 – Yes). Data reported by TIMSS countries coordinators and available in the TIMSS 2019 Encyclopedia.

Data analysis

Students, schools and science teachers data files were merged with the intsvy (Caro & Biecek 2017, v.2.9) for the R statistical system (R Core Team 2024, v. 4.3). Science country means and standard errors were estimated considering the five plausible values (*PV*) for Science, using the appropriate estimators for the complex design of TIMSS 2019, with the intsvy package. The multicollinearity of the explanatory variables was checked with the Variance Inflation Factor (*VIF*) from the package car (Fox & Weisberg 2019, v. 3.0.13). No multicollinearity problems were assumed for $VIF<5$.

A three-level hierarchical linear model was fitted to the five science PV and student (level 1), school/teachers (level 2), and countries (level 3) variables using the package WeMix (v. 4.0 Bailey, et al., 2023) with the additional function mixPV (Huang 2024). mixPV allows for the automatic pooling of all PV with the proper estimation of HLM coefficients, robust standard errors, and degrees of freedom as required for modeling with plausible values (Rutkowski et al. 2010).

Level 2 clusters were created by concatenating the country ID (IDCNTRY) and school ID (IDSCHOOL). Level 3 clusters were the countries with schools nested within countries and students nested within schools. Student level and school/teachers level were weighted by Senate weights and school weights, respectively, present in the TIMSS 2019 data bank. These unconditional Level 1 and Level 2 weights were normalised by the overall mean of the original senate and school weights, respectively. All country weights were set at 1 so that all countries contributed equally to the analysis. These weights were designed to take into account both the sampling design and adjustments for non-responses.

A three-level HLM random intercepts, constant slopes model was fitted to the data in four steps. In the first step, all student-level $\mathbf{X}_{ijk}$ variables for all the i students nested in the j schools and k countries were probed:

$$PV_{ijk}^{1-5} = \beta_{0jk} + \beta_{1jk}\mathbf{X}_{ijk} + \varepsilon_{ijk}.$$

In the second step, the $\mathbf{W}_{jk}$ level 2 schools/teachers variables were probed for significance as follows:

$$\boldsymbol{\beta}_{0jk} = \gamma_{00k} + \gamma_{01k}\mathbf{W}_{jk} + u_{0jk}.$$

In the third step, the level 3 country $\mathbf{Z}_k$ variables were probed for significance as follows:

$$\boldsymbol{\gamma}_{00k} = \pi_{000} + \boldsymbol{\pi}_{001}\mathbf{Z}_k + v_{00k}.$$

At the final step, all significant variables in steps 1, 2, and 3 were combined. We avoided the more traditional nested testing of HLM (adding level 2 variables to the model with selected level 1 variables, and adding level 3 variables, to selected level 2 and level 1 variables) to overcome the missing data on several level 1 variables (see variables description above). Intraclass correlation coefficients were calculated as follows:

$$ICC_{school} = \frac{\sigma^2_{u_{0jk}}}{\sigma^2_{u_{0jk}} + \sigma^2_{v_{00k}} + \sigma^2_{\varepsilon_{ijk}}},$$

$$ICC_{country} = \frac{\sigma^2_{v_{00k}}}{\sigma^2_{u_{0jk}} + \sigma^2_{v_{00k}} + \sigma^2_{\varepsilon_{ijk}}}.$$

Marginal R^2, that is the amount of total variance explained by the fixed effects, was calculated as

$$R^2_m = \frac{\sigma^2_f}{\sigma^2_f + \sigma^2_{u_{0jk}} + \sigma^2_{\gamma_{00k}} + \sigma^2_{\varepsilon_{ijk}}}.$$

The conditional R^2, that is the amount of total variance explained by the model (fixed + random effects), was calculated as follows:

$$R^2_c = \frac{\sigma^2_f + \sigma^2_{u_{0jk}} + \sigma^2_{\gamma_{00k}}}{\sigma^2_f + \sigma^2_{u_{0jk}} + \sigma^2_{\gamma_{00k}} + \sigma^2_{\varepsilon_{ijk}}},$$

where

$$\sigma^2_f = \text{var}\left(\sum_{h=1}^{p} \beta_h \mathbf{X}_{hijk}\right) + \text{var}\left(\sum_{h=1}^{s} \gamma_h \mathbf{W}_{hjk}\right) + \text{var}\left(\sum_{h=1}^{c} \pi_h \mathbf{Z}_k\right)$$

is the variance of the fixed effects, $\sigma^2_{u_{0jk}}$ plus $\sigma^2_{\gamma_{00k}}$ is the variance of the random effects, and $\sigma^2_{\varepsilon_{ijk}}$ is the residual variance (Nakagawa & Schielzeth 2013).

Standardised beta coefficients (β) were used as measures of effect sizes, and statistical significance was accepted for p<.05. A β value of 0.10 to 0.19 is considered a small effect, 0.20 to 0.29 is a medium effect, and 0.30 or greater is a large effect (Cohen 1998; Nieminen 2022).

Results

Grade 8 science literacy around the world, as estimated by the replicated means of the 5 science plausible values, is displayed in Figure 1. South Africa (M = 372, SE = 3.18), Lebanon (M = 377, SE = 4.65), Egypt (M = 389, SE = 5.44), and Morocco (M = 394.1, SE = 2.66), all African countries, achieved the lowest science scores in the TIMSS 2019 edition. Singapore (M = 608, SE = 3.91), Chinese Taipei (M = 575, SE = 1.98), and Japan (M = 571, SE = 2.63), all Asian countries, were top performers. Most European countries had scores around the TIMSS mean (e.g, France M = 488, SE = 2.89; Norway M = 496, SE = 3.27; Italy M = 500, SE = 2.57; England M = 517, SE = 4.82; Portugal M = 519, SE = 2.72 and Finland M = 542, SE = 3.24). In North America, only the USA participated (M = 522, SE = 4.73). Australia (M = 527, SE = 3.46) and New Zealand (M = 500, SE = 3.48) represented Oceania.

Science mean scores differences around the world were strongly associated with the UNDP Human Development Index (r = 0.79, p<.001) but not with the countries' total expenditure in education (as % of the GDP) (r = 0.03, p = 0.86) (see Figure 2). Countries

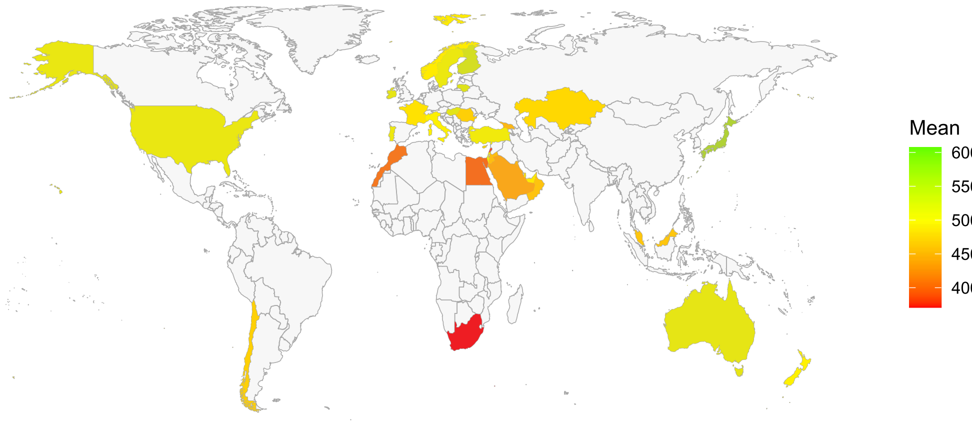


Figure 1. Mean grade 8 science literacy from countries who participated in the TIMSS 2019 grade 8 edition. The TIMSS scale has a mean of 500 points with a standard deviation of 100 points.

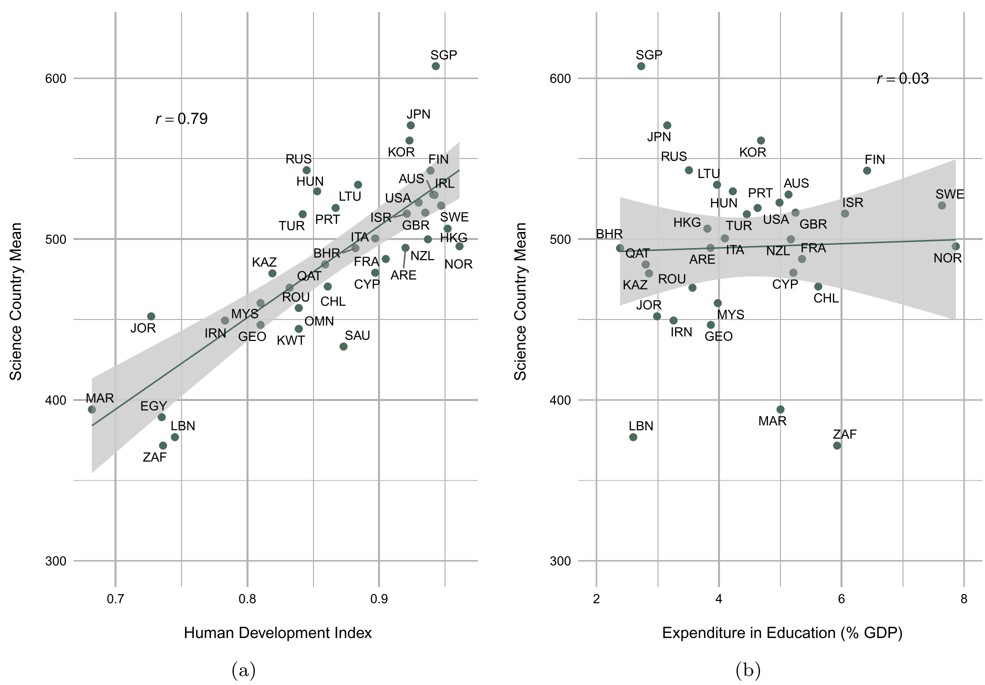
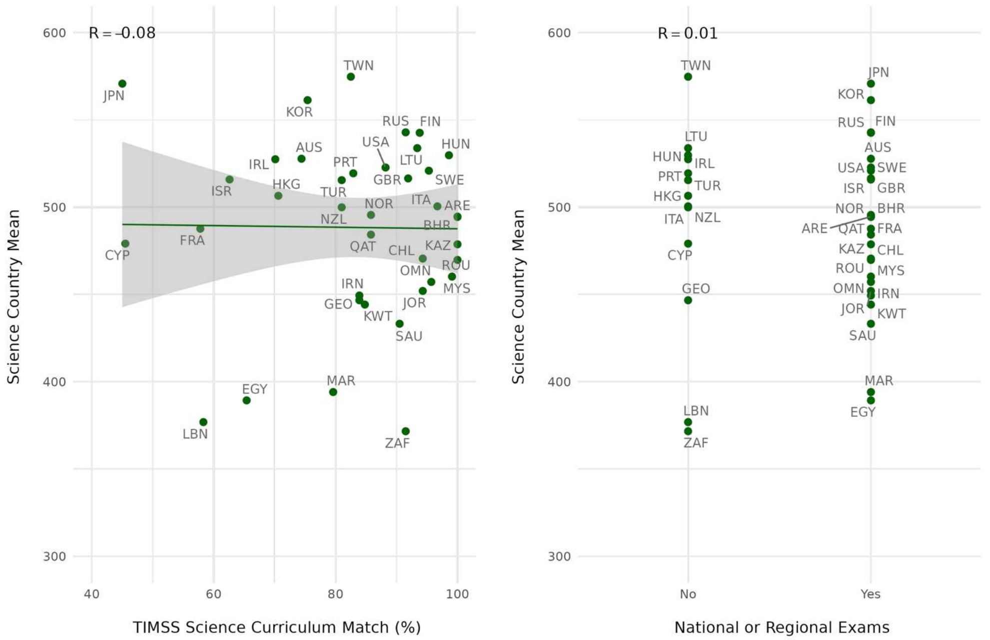


Figure 2. TIMSS 2019 Science mean scores at grade 8 by country as a function of UNDP Human Development Index (a) and Total Expenditure in Education (as % of GDP) (b). R is the Pearson correlation coefficient; $r = 0.79$; $p<.001$ for Science Country mean score vs. Human Development Index; and $r = 0.05$, $p = 0.80$ for Science Country mean score vs. Expenditure. Gray bands are the 95% confidence intervals for the regression lines.

Note: ARE – United Arab Emirates, AUS – Australia, BHR – Bahrain, CHL – Chile, CYP – Cyprus, EGY – Egypt, FIN – Finland, FRA – France, GBR – United Kingdom, GEO – Georgia, HKG – Hong Kong SAR China, HUN – Hungary, IRL – Ireland, IRN – Iran, ISR – Israel, ITA – Italy, JOR – Jordan, JPN – Japan, KAZ – Kazakhstan, KOR – South Korea, KWT – Kuwait, LBN – Lebanon, LTU – Lithuania, MAR – Morocco, MYS – Malaysia, NOR – Norway, NZL – New Zealand, OMN – Oman, PRT – Portugal, QAT – Qatar, ROU – Romania, RUS – Russia, SAU – Saudi Arabia, SGP – Singapore, SWE – Sweden, TUR – Turkey, TWN – Taiwan, USA – United States, ZAF – South Africa.



(a) (b)

Figure 3. Relationship between the National Curriculum match with the TIMSS Science grade 8 assessed curriculum (% of items in the TIMSS test covered by national curricula) and Science Mean Scores (a) and the existence of national or regional exams and the countries means (b). R is the Pearson correlation coefficient (a) or the Biserial Correlation (b). Grays bands are the 95% confidence intervals for the regression lines.

with higher human development scored higher in TIMSS 2019 Science for grade 8. This was particularly the case for Japan and South Korea in Asia, and Finland in Europe. On the opposite hand, African countries like Morocco or South Africa had the lowest mean scores (see Figure 2(a)). Contrary to popular belief, countries with higher total expenditure in Education (as % of GDP), like Sweden or Norway, did not perform significantly better than countries with lower investment, like Bahrain or Qatar (see Figure 2(b)). Countries' coverage of the Science TIMSS curriculum was also not associated with countries science literacy (see Figure 3(a)) but Schools' Science Self-evaluation was associated with Science Literacy (see Figure 3(b)). These results partially confirm our H_1 since only HDI was strongly and positively associated with science literacy at grade 8.

As per the basal Hierarchical Linear Model, Countries' differences accounted for 25.8% of the total variation in TIMSS 2019 8th grade science scores around the world ($ICC_{Country} = 0.258$), while schools within countries accounted for 22.8% of the total variation in science achievement ($ICC_{School} = 0.228$) (see 'Basal Model', Table 1). To identify which students, teachers/schools, and countries variables could be associated with science literacy at grade 8, we performed a 3-level (students, teachers/schools, and country) Hierarchical Linear Modeling. Explanatory variables were checked for multicollinearity before entering the HLM (VIFs ranged from 1.02 for BSBG01 to 2.9 for BSBGSLS). Table 1 shows the standardised beta coefficients (β) for level 1 (students) variables, level 2 (Teachers and Schools) variables and level 3 (countries) variables added

Table 1. Fixed-effects standardised regression coefficients (β) with *SE* between parentheses for the student-level variables, school and teacher level, and country level, and random effects variance.

Variables	Basal	Std. level	Sch. level	Country level	Selected variables
Fixed-effects β(*SE*)					
(Intercept)	-0.003 (0.080)	0.238 (0.105)*	0.018 (0.070)	-0.037 (0.051)	0.064 (0.106)
BSBG01		0.001 (0.011)			
BSBGHER		0.179 (0.020)***			0.167 (0.016)***
BSBGSSB		-0.010 (0.013)			
BSBGSB		0.048 (0.014)***			0.043 (0.015)**
BSBGSLS		0.057 (0.013)***			0.067 (0.019)***
BSBGICS		-0.040 (0.008)***			-0.040 (0.010)***
BSBGSCS		0.199 (0.022)***			0.200 (0.024)***
BSBGSVS		0.016 (0.019)			
BSBGSEC		0.074 (0.010)***			0.075 (0.010)***
BCBG03A			-0.113 (0.031)***		-0.120 (0.024)***
BCBG06A			0.005 (0.018)		
BCBG08A			0.049 (0.015)**		0.041 (0.012)***
BCBG10A			0.034 (0.012)**		0.015 (0.008)+
BCBGDAS			0.073 (0.013)***		0.033 (0.010)**
BCBGEAS			0.140 (0.016)***		0.081 (0.017)***
BCBGSRS			0.018 (0.014)		
BTBG01			0.009 (0.004)*		-0.009 (0.016)
BTBG04			0.011 (0.004)**		0.022 (0.008)*
BTBG05G			0.001 (0.004)		
BTBGLSN			0.051 (0.016)***		0.113 (0.041)**
BTBGTJS			-0.004 (0.004)		
BTBSESI			0.010 (0.004)**		0.012 (0.015)
HDI				0.414 (0.061)***	0.253 (0.143)+
ExpEdu				-0.098 (0.044)*	0.013 (0.161)
PercCurrMatch				-0.022 (0.060)	
NatRegExams				0.077 (0.050)	
SchSelfEvaluation				0.054 (0.022)*	0.152 (0.104)
Random Effects Var(SE)					
School.(Intercept)	0.203 (0.023)***	0.118 (0.020)***	0.137 (0.098)	0.198 (0.024)***	0.081 (0.015)***
Country.(Intercept)	0.229 (0.059)***	0.116 (0.061)+	0.178 (1.647)	0.068 (0.017)***	0.090 (0.043)*
Residual	0.457 (0.018)***	0.391 (0.021)***	0.452 (0.019)***	0.447 (0.018)***	0.386 (0.022)***
Statistics					
ICC_{School}	0.228	0.133	0.154	0.223	0.091
$ICC_{Country}$	0.258	0.130	0.200	0.076	0.101
R^2_m	0.000	0.162	0.127	0.161	0.367
R^2_c	0.486	0.476	0.485	0.474	0.562
n_{obs}	354959	80459	257637	326481	64662
n_{PV}	5	5	5	5	5
$\bar{AIC}$	738888.9	128477.7	544681.4	677111.9	106603.5
$\bar{BIC}$	738932.0	128598.5	544859.2	677208.1	106803.2

$^{***}p \leq .001$; $^{**}0.001 < p \leq .01$; $^{*}0.01 < p \leq 0.05$; $^{+}0.05 < p \leq 0.10$.

sequentially to the three-level HLM. All student-related variables, except for Sex (BSBG01), Sense of School Belonging (BSBGSSB) and Students valuing science (BSBGSVS), were statistically significant (p<.005). However, only the Home Resources for Learning (BSBGHER, a proxy of socioeconomic and cultural status of students) and Students' Self-confidence in Science (BSBGSCS) displayed non-trivial effect sizes (Standardised regression slope $\beta > 0.1$). Students with higher scores on the Home Resources for Learning scale displayed higher science scores ($\beta = 0.179$; $SE = 0.020; p < .001$). In the same direction, students with higher self-confidence in Science obtained higher scores in Science literacy ($\beta = 0.199$; $SE = 0.020$; $p < .001$). Student-level variables resulted in a reduction of 14% of the residual (students) variance. The fixed effects explained 16.2% of the total variance ($R_m^2 = 0.162$) and the conditional model explained 47.6% of the total science literacy variance ($R_c^2 = 0.476$). It did not escape our attention that there was a considerable missing data proportion in student variables (78%) because 12 countries did not apply one or more of the students scales. This may explain why some student attitudes (e.g. students valuing science) did not display a significant effect in the model. These results confirm our H_2: students' science literacy is positively associated with students socioeconomic resources and students' motivation and self-confidence.

In the second step of the analysis, school and teacher (aggregated) variables were added to the model. Only the percentage of economically disadvantaged students attending school (BCBG03A) and the Schools emphasis on academic success (BCBGEAS) displayed statistically significant effects with non-trivial size. Schools with higher percentages of economically disadvantaged students had lower science literacy scores ($\beta = -0.113$; $SE = 0.031$; $p < .001$). Schools with higher emphasis in academic success displayed better science literacy ($\beta = 0.140$; $SE = 0.016$; $p < .001$). Although several teachers variables displayed statistical significance, their effect sizes were trivial (e.g. BTBG01-Years been teaching with $\beta = 0.009$; $SE = 0.004$; $p = 0.049$; or BTBSESI-Teachers emphasis in science investigations with $\beta = 0.010$; $SE = 0.004; p = 0.008$). School and Teachers variables reduced the variation between schools by 33% and their fixed effects explained 12.7% ($R_m^2 = 0.127$) of the total variance while the whole model explained 48.5% ($R_c^2 = 0.485$). Our H_3 was partially confirmed, as only the school's percentage of economically disadvantaged students and the school's emphasis on academic success were found to be associated with the students' science literacy, exhibiting non-trivial effect sizes. Although statistically significant, the teachers variables displayed relatively small effect sizes.

At the country level, only the UNDP Human Development Index (HDI) was strongly and positively associated with science literacy ($\beta = 0.414$; $SE = 0.061$; $p < .001$). Not surprisingly, because this trend was already seen in Figure 3(a). The percentage of TIMSS science items that were covered by the national curricula (PercCurrMatch), after accounting for other country variables, was weakly but negatively associated with science scores ($\beta = -0.098$; $SE = 0.061$; $p < .001$). Schools that promoted science self-evaluations (SchSelfEvaluation) displayed better science scores ($\beta = 0.054$;$SE = 0.022$; $p = 0.046$) but again the effect size was trivial. Adding country-level variables reduced the between-country variance by 70%. The country's fixed effects explained 16.1% of the total science literacy variance ($R_m^2 = 0.161$). These results reinforce the partial support of H_1 as described before.

In the last step of the HLM analysis, we combined all the statistically significant variables from the previous steps (see column 'Selected variables' in Table 1). At the student level, home resources for learning and students self-confidence in science were the only ones with statistically significant non-trivial effect sizes ($\beta = 0.167$; $SE = 0.016$; $p < .001$, and $\beta = 0.200$; $SE = 0.024$; $p < .001$, respectively). At the school and teachers level, the percentage of economically disadvantaged students kept is statistically significant and has a non-trivial negative effect size ($\beta = -0.112$; $SE = 0.024$; $p < .001$) as well and the emphasis on academic success ($\beta = 0.081$; $SE = 0.017$; $p < .001$). However, after accounting for the students' and country-level variables effects, the teachers reported teaching limited by students' needs acquired a non-trivial and statistically significant effect size ($\beta = 0.113$; $SE = 0.041$, $p < .001$). The lower the teaching limitations due to student needs the better the science literacy. Finally, at the country level, after accounting for students and schools/teachers variables, HDI loses strength ($\beta = 0.167$; $SE = 0.016$; $p = 0.014$). The overall model reduces the school-related variance by 60%, the country-related variance by 61%, and the student-related variance by 16%. The combined fixed effects (students, schools/teachers, and countries) explained 36.7% of the total variance ($R^2_m = 0.367$), and the combined fixed and random effects explained 56.2% of the total science literacy variance ($R^2_c = 0.562$).

Discussion

Science education is fundamental for modern societies as it equips individuals with the knowledge and skills necessary to understand and navigate the complexities of the world around them. In an era marked by rapid technological advancements and complex global challenges such as climate change, pandemics, and resource depletion, a strong foundation in science is indispensable. It empowers citizens to make informed decisions, critically evaluate information, and participate meaningfully in civic discourse. Moreover, science education fosters curiosity, creativity, and problem-solving abilities, laying the groundwork for innovation and progress. IEA's TIMSS is an ongoing research effort that evaluates science literacy both at grade four, as well as grade eight students from all over the world. Here, we probed students, school and teachers, and country-level variables that are associated with grade eight students' science literacy.

Students and families effects

At the student level, the influence of home resources for learning and students' self-confidence in science stands out as primary factors in explaining science literacy, serving as proxies for family socioeconomic and cultural status. This finding resonates with numerous studies across various literacies. For instance, Marôco (2021) illustrates this correlation in the context of reading literacy, while Kampman et al. (2022) demonstrate its relevance in Math and Science Literacies.

Teachers and schools

Economic factors, such as the percentage of economically disadvantaged students, emerge as significant negative correlates of science literacy at the school level. Similar results have been observed elsewhere. For example, in Australia, Perry and McConney (2010) observed that increases in the mean SES of a school are associated with consistent increases in students academic achievement and that this relationship is similar for all students regardless of their individual SES. However, at the country level, the percentage of GDP allocated to education does not emerge as a predictor of science literacy. Instead, the UNDP Index of Human Development, accounting for individuals variables like life expectancy, education levels, and income per capita, emerges as the strongest predictor of science literacy. This is consistent with the effects of higher teaching limitations due to students' needs in reduced achievement. These results underscore the pivotal role of family economic status, rather than the countries investment in education for successful science learning among eighth-grade students. Our observations are consistent with (Kirkcaldy et al. 2006) observations, with OECD's PISA data in 30 nations, that GDP and economic growth were very weakly related to educational performance of 15 years old on the PISA test. On the other hand, inflation and the human development index were significantly related to all three PISA literacy scores (Kirkcaldy et al. 2006).

Countries and educational systems

Notably, characteristics of education systems, such as the alignment of countries' science curricula with the curriculum assessed by TIMSS or the existence of high-stakes exams, exhibit a small or no statistical significance in predicting science literacy once the effects of students, schools, and teachers variables are considered. International tests like TIMSS and PISA are influencing how countries perceive science education. In 'Making it Comparable: Standards in Science Education' (Waddington et al. 2007), focusing on science standards development, particularly in northern Europe, half of the countries referenced TIMSS or PISA results in discussing their approach to curriculum changes. Consequently, many nations are refining student outcome statements to enhance science learning and test scores, reflecting a trend towards a globalized science education (DeBoer 2011). However, our analysis indicates that alignment with the TIMSS-assessed curriculum is not a significant predictor of a country's science literacy at grade 8. This suggests that while curriculum alignment and assessment may be relevant in some contexts, individual and contextual factors play a larger role in shaping science literacy outcomes. On the other hand, our study shows that it is not merely the amount of money spent on science education, measured as a percentage of GDP, that leads to improvements in science literacy among grade 8 students. Instead, a more comprehensive investment in overall human development proves to be more effective (Hanushek & Woessmann 2011). This includes investments in income, education, and health, which together create a supportive environment for students. A higher income level can reduce economic stress on families, allowing students to focus more on their studies (UNDP 2019). Improved education systems ensure better quality teaching, access to learning resources, and curriculum development that fosters critical thinking and problem-solving skills. Investments in health, including nutrition, mental health, and physical well-being, are crucial as healthy students are more likely to perform better academically (Filmer & Pritchett 1999). These factors combined contribute significantly to creating an environment where students can thrive and achieve higher levels of science literacy. This holistic approach to development underscores the importance of addressing multiple aspects of a child's environment, rather than focusing solely on direct educational expenditures.

Policy implications

The findings from this study suggest several important policy implications that can enhance science literacy among grade eight students. Firstly, targeted interventions to support socioeconomically disadvantaged students are crucial, as family economic status significantly impacts science literacy. Policymakers should consider programmes that provide additional learning resources, tutoring, and support mechanisms, such as after-school programmes and summer schools, to help bridge the gap in science literacy and motivation for these students. Secondly, enhancing teacher training and support, particularly in science education, is essential for improving instructional quality and student outcomes. Continuous professional development programmes focused on the latest scientific pedagogies, classroom management techniques, and the integration of technology in teaching can empower teachers to deliver more effective science instruction, engage students and promote science learning.

Furthermore, curriculum development should be comprehensive and contextualised, rather than merely aligning with international assessments like TIMSS. Policymakers should engage educators in the curriculum development process to ensure that the curriculum is relevant to the local context and meets the specific learning needs and cultural backgrounds of students. Emphasizing inquiry-based learning, critical thinking, and real-world applications of science can make the curriculum more engaging and effective.

Policymakers should focus on holistic development indicators such as the Human Development Index (HDI), which encompasses life expectancy, education levels, and income per capita, as these are stronger predictors of science literacy than GDP allocation to education. This holistic approach ensures that broader developmental goals are met, creating a more supportive environment for educational achievement. International collaboration and knowledge sharing can also facilitate the exchange of effective educational practises. Countries can learn from each other's successes and challenges, fostering a global improvement in science education through collaborative efforts and shared resources.

Addressing students' motivation for low-stakes tests is another critical area. Integrating these assessments into regular grading systems or providing incentives for good performance could enhance student engagement and effort. Educating students about the importance of these assessments for informing educational policy and practice can also improve their motivation and performance. Lastly, ongoing research and comprehensive data collection are vital to better understand the multifaceted factors influencing science literacy around the world. Policymakers should support initiatives, like TIMSS, that ensure robust data collection across all relevant scales and contexts, enabling more accurate analysis and more effective educational interventions. By implementing these strategies, policymakers can create a more scientifically literate population, better equipped to navigate and address the complex challenges of the modern world.

Limitations

These results are not devoid of methodological limitations. Firstly, at the student level, there are concerns regarding the extent of effort students allocate to these types of low-stakes tests. While there are no reports for TIMSS, the OECD's Programme for International Student Assessment (PISA) evaluates 15-year-old students in Science, Math, and Reading literacies through this type of inquiry. In PISA 2018, students were questioned about the effort they exerted on the test and how much effort they would have expended if the test results were factored into their grades. Across OECD countries, 68% of students reported exerting less effort on the PISA test compared to a test that influenced their grades (OECD 2019). Thus, concerns arise regarding the efficacy of international large-scale assessments (ILSA) like TIMSS or PISA in evaluating students' knowledge. To date, only a few published reports have examined the concurrent validity of ILSA and national high-stakes exams. For instance, Marôco (2020) identified moderate to high correlations between TIMSS 2015 math scores for grades 4 and 12 and national exams at these grades ($R = 0.71$, $p<.001$), as well as between PISA 2015 math scores and the national math exam at grade 9 ($R = 0.63$;$p<.001$). In Finland, Pulkkinen & Rautopuro (2022) observed that PISA proficiency scores correlated not only with the corresponding grades but also with the grades of other theoretical subjects, indicating that the PISA test assesses a wide range of school achievement. While more information from other countries is needed, these two reports support the concurrent validity of ILSA with national assessments. Additionally, 12 out of 38 participating countries and regions did not administer several student-related scales (e.g. Students Like Learning Science), which restricted hierarchical linear modeling (HLM) estimation of these variables' effects and the available sample size for modeling. This may elucidate the minimal effect sizes observed for some variables and constrain the generalisation of the conclusions.

Concluding remarks

As highlighted by the ongoing research efforts of IEA's TIMSS, understanding the factors that influence science literacy at the eighth-grade level is paramount. At the heart of these influences lie not only economic factors but also the dynamics of home resources for learning and students' self-confidence in science. These factors, reflective of family socio-economic and cultural status, play a pivotal role in shaping science literacy outcomes. Importantly, investment in education at the national level does seem to be a key element as it is overshadowed by the broader indices of human development. This emphasises the multifaceted nature of the international science education landscape. Moreover, the limited predictive power of education system characteristics highlights the need to focus on individual and contextual factors in fostering science literacy. Ultimately, by elucidating these dynamics, we can better tailor educational interventions to empower students with the necessary tools to engage meaningfully with science and contribute to innovation and progress in our society.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

João Marôco http://orcid.org/0000-0001-9214-5378
Heidi Harju-Lukkainnen http://orcid.org/0000-0002-4532-7133
Juhani Rautopuro http://orcid.org/0000-0002-3231-4843

Ethic statement

Ethics approval was not required since the data are freely available.

References

Bailey, P., Webb, B., Kelley, C., Nguyen, T., & Huo, H. (2023). WeMix: Weighted mixed-effects models using multilevel pseudo maximum likelihood estimation (Version 4.0.0) [R package]. CRAN.

Caro, D. H., & Biecek, P. (2017). intsvy: An R package for analyzing international large-scale assessment data. *Journal of Statistical Software*, *81*(7), 1–44. https://doi.org/10.18637/jss.v081.i07

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

DeBoer, G. E. (2011). The globalization of science education. *Journal of Research in Science Teaching*, *48*(6), 567–591. https://doi.org/10.1002/tea.v48.6

Dickman, A., Schwabe, A., Schmidt, J., & Henken, R. (2009). *Preparing the future workforce: Science, technology, engineering, and math (STEM) policy in K-12 education*. Public Policy Forum.

Ersan, O., & Rodriguez, M. C. (2020). Socioeconomic status and beyond: A multilevel analysis of TIMSS mathematics achievement given student and school context in Turkey. *Large-scale Assessments in Education*, *8*, 15. https://doi.org/10.1186/s40536-020-00093-y

Filmer, D., & Pritchett, L. (1999). The impact of public spending on health: Does money matter? *Social Science & Medicine*, *49*(10), 1309–1323. https://doi.org/10.1016/S0277-9536(99)00150-1

Fischman, G. E., Topper, A. M., Silova, I., Goebel, J., & Holloway, J. L. (2019). Examining the influence of international large-scale assessments on national education policies. *Journal of Education Policy*, *34*(4), 470–499. https://doi.org/10.1080/02680939.2018.1460493

Fishbein, B., Foy, P., & Yin, L. (2021). TIMSS 2019 international database. In *TIMSS 2019 user guide for the international database* (2nd ed.). TIMSS & PIRLS International Study Center.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage. Available at https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Foy, P., Fishbein, B., von Davier, M., & Yin, L. (2020). Implementing the TIMSS 2019 scaling methodology. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (pp. 12.1–12.146). TIMSS & PIRLS International Study Center.

Hanushek, E. A., & Woessmann, L. (2011). The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 89–200). North-Holland.

Hodson, D. (2003). Time for action: Science education for an alternative future. *International Journal of Science Education*, *25*(6), 645–670. https://doi.org/10.1080/09500690305021

Huang, F. L. (2024). Using plausible values when fitting multilevel models with large-scale assessment data using R. *Large-scale Assessments in Education*, *12*, Article 7.

IEA (2024). Trends in international math and science study (TIMSS).

Johansson, S. (2016). International large-scale assessments: What uses, what consequences?. *Educational Research*, *58*(2), 139–148. https://doi.org/10.1080/00131881.2016.1165559

Johansson, S., & Yang Hansen, K. (2019). Are mathematics curricula harmonizing globally over time? evidence from TIMSS national research coordinator data. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*(2), em1656.

Kampman, K., Geske, A., & Ozola, A. (2022). The influence of family socioeconomic status on students self-beliefs in large scales studies. *Human, Technologies and Quality of Education*.

Kaya, M. (2022). The relationship among socioeconomic status, attitude towards science, school climate and students' science achievement: A cross-country comparison of TIMSS. *Science Education International*, *33*(4), 366–375. https://doi.org/10.33828/sei

Kirkcaldy, B., Furnham, A., & Siefen, G. (2006). The relationship between health efficacy, educational attainment, and well-being among 30 nations. *European Psychologist*, *9*(2), 107–119. https://doi.org/10.1027/1016-9040.9.2.107

Maôco, J. (2020). International large-scale assessments: Trends and effects on the Portuguese public education system. In H. Harju-Luukkainen, N. McElvany, & J. Stang (Eds.), *Monitoring student achievement in the 21st century*. Springer. https://doi.org/10.1007/978-3-030-38969-7\17
Marôco, J. (2021). What makes a good reader? Worldwide insights from PIRLS 2016. *Reading and Writing: An Interdisciplinary Journal*, *34*(1), 231–272. https://doi.org/10.1007/s11145-020-10068-8
Martin, M. O., von Davier, M., & Mullis, I. V. S. (2020). *Methods and procedures: TIMSS 2019 technical report*. TIMSS & PIRLS International Study Center.
Mullis, I. V. S., & Martin, M. O. (2017). *TIMSS 2019 assessment frameworks*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/frameworks/.
Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/mee3.2013.4.issue-2
Nieminen, P. (2022). Application of standardized regression coefficient in meta-analysis. *BioMedInformatics*, *2*(3), 434–458. https://doi.org/10.3390/biomedinformatics2030028
OECD (1996). The knowledge-based economy. OECD Publishing.
OECD (2019). PISA 2018 results (Volume I): What students know and can do. OECD Publishing.
OECD (2023). PISA 2022 assessment and analytical framework. OECD Publishing.
Perry, L. B., & McConney, A. (2010). Does the SES of the school matter? an examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, *112*(4), 1137–1162. https://doi.org/10.1177/016146811011200401
Pulkkinen, J., & Rautopuro, J. (2022). The correspondence between PISA performance and school achievement in Finland. *International Journal of Educational Research*, *114*, Article 102000. https://doi.org/10.1016/j.ijer.2022.102000
R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing.
Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, *39*(2), 142–151. https://doi.org/10.3102/0013189X10363170
UNDP (2019). Human development report 2019: Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century. UNDP.
UNESCO (1999). Declaration on science and the use of scientific knowledge.
van Driel, J. H., Meirink, J. A., van Veen, K., & Zwart, R. C. (2012). Current trends and missing links in studies on teacher professional development in science education: A review of design features and quality of research. *Studies in Science Education*, *48*(2), 129–160. https://doi.org/10.1080/03057267.2012.738020
von Davier, M. (2020). TIMSS 2019 scaling methodology: Item response theory, population models, and linking across modes. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (pp. 11.1-11.25). TIMSS & PIRLS International Study Center.
Waddington, D., Nentwig, P., & Schanze, S. (2007). *Making it comparable: Standards in science education*. Waxmann.
Yin, L., & Fishbein, B. (2020). Creating and interpreting the TIMSS 2019 context questionnaire scales. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report*. TIMSS & PIRLS International Study Center.