Bilingualism: Language and Cognition

cambridge.org/bil

Unlocking mathematical potential through school-based language learning: Insights from PISA 2018

Alejandra Nucette[1], Britta Biedermann[1,2], Suze Leitão[1] and Takeshi Hamamura[3]

[1]School of Allied Health, Curtin University, Australia; [2]EnAble Institute, Curtin University, Australia and [3]School of Population Health, Curtin University, Australia

Research Article

Cite this article: Nucette, A., Biedermann, B., Leitão, S. and Hamamura, T. (2026). Unlocking mathematical potential through school-based language learning: Insights from PISA 2018. *Bilingualism: Language and Cognition* 1–17. https://doi.org/10.1017/S1366728926101138

Received: 30 July 2025
Revised: 16 January 2026
Accepted: 9 February 2026

Keywords:
foreign language learning; mathematics achievement; multicultural curriculum; adolescents; educational equity

Corresponding author:
Alejandra Nucette;
Email: alejandra.nucette@postgrad.curtin.edu.au

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

This study explores the association between school-based foreign language (FL) instruction and mathematical achievement among 15-year-old students, using data from the 2018 Programme for International Student Assessment (PISA). Two complementary analyses were conducted: a large-scale model (n = 300,656) examining the relationship between time spent in FL learning and maths performance across 73 countries and a machine learning (ML) approach (random forest (RF); n = 53,459) identifying specific programme features that most strongly influence this relationship. Results show that longer exposure to FL instruction was associated with a modest but statistically robust increase in maths scores (β = 0.08, p < .001), even after controlling for socioeconomic and contextual factors. Among programme characteristics, the integration of multicultural curricula emerged as a prominent predictor of higher maths performance. These findings indicate that sustained, culturally enriched FL learning is positively associated with numeracy outcomes, with implications for equity in academic achievement and cross-disciplinary performance.

1. Introduction

Despite its potential significance, the association between foreign language (FL) learning and maths performance remains an underexplored area of research. This study examines how acquiring a new language may do more than broaden linguistic horizons; it explores the hypothesis that FL instruction shares critical cognitive pathways with mathematical reasoning, particularly during the pivotal stage of adolescence.

Using data from over 300,000 students across 73 countries in the 2018 Programme for International Student Assessment (PISA), our research addresses the question: How is participation in language learning associated with mathematics performance? By identifying a substantial predictive association, this research suggests that FL programmes warrant consideration as potential strategic markers for global academic achievement and educational equity.

1.1. Foreign language learning and mathematical achievement: A cognitive perspective

Learning an FL for the first time – including in mainstream school settings – has been theorised to enhance executive function (EF), particularly working memory, cognitive flexibility, and inhibitory control. These EF components are strongly associated with academic success, especially maths performance (Fuhs et al., 2014). Based on these findings, researchers have hypothesised that an indirect link between FL learning and mathematical achievement may exist, mediated by improvements in EF, and potentially shaped by instructional factors (Schiltz et al., 2024).

Although this link has been previously explored, the body of research remains relatively small and has primarily focused on younger cohorts. Among low-income preschoolers, bilingualism has been associated with mathematics through growth in inhibitory control (Choi et al., 2018). In primary school, working memory gains due to FL learning predict mathematical problem-solving skills among language learners in grades 1–4 (Lee Swanson et al., 2021). Similarly, in two-way dual-language programmes, gains in mathematical skills among fourth and fifth graders were statistically mediated by inhibitory control, working memory, and cognitive flexibility – an effect not observed in younger groups (Esposito, 2020). Consistent with this pattern, Park et al. (2023) reported that enhanced EF associated with FL learning is linked to higher mathematics scores in fourth graders, after adjusting for caregiver education.

Research on the cognitive benefits of FL learning initiated during adolescence is also limited. However, evidence from later life stages suggests that these extend well beyond early childhood. Working memory improvements following FL learning are well-documented, with significant gains observed in both young and older adult learners after up to a year of FL instruction (Huang

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

et al., 2022; Wong et al., 2019). Importantly, the literature shows that adolescents (average age 16 years) have also shown measurable improvements (Shoghi & Ghonsooly, 2018).

FL learning after childhood also strengthens cognitive flexibility and attention switching. While few studies focus exclusively on adolescents, research demonstrates that even short-term instruction results in benefits for learners aged 16–78 (Al-khresheh & Karmi, 2024; Bak et al., 2016; Shoghi & Ghonsooly, 2018; Vega-Mendoza et al., 2015). Likewise, inhibitory control improves through FL learning, with significant gains observed in children, adolescents, and young adults after as little as six months of instruction (Bialystok & Barac, 2012; Sullivan et al., 2014).

These cognitive skills – working memory, flexibility, and inhibition – are critical for managing complex tasks and understanding mathematical concepts across developmental stages (Bull & Lee, 2014; Cragg et al., 2017; ten Braak et al., 2022). Conversely, deficits in these skills are linked to mathematical learning difficulties (Iglesias-Sarmiento et al., 2023; Ramos et al., 2018), which has prompted the use of EF-targeted interventions (Sabine, 2020). Taken together, these findings support the rationale for investigating whether FL learning beginning during adolescence can yield cognitive benefits, with potential implications for mathematical achievement.

Beyond cognitive theory, empirical evidence suggests a positive link between FL learning and higher mathematics scores. However, most existing studies are geographically concentrated in the United States and Canada. Research on primary school dual-language immersion programmes indicates that students outperform their nonlanguage-learning peers in mathematics, even after adjusting for socioeconomic status (SES) and ethnicity (Padilla et al., 2013; Watzinger-Tharp et al., 2016). Similarly, Lee (2010), in their study of South Korean Year 11 students, identified a strong correlation between proficiency in English as a second language and mathematics achievement across both general and advanced maths courses.

Additionally, Woll and Wei's (2019) systematic review of 20 studies examining preschoolers to university students found that 90% of the included research reported positive correlations between FL learning and academic outcomes, including mathematics. Moreover, a recent meta-analysis involving over 785,000 students (aged 10–17) found that second-language learners are three times more likely to achieve higher math grades compared to their non-learner peers (Nucette et al., 2024). These results were consistent across various FL delivery methods, including dual-immersion, FL, and English as a second-language programmes.

At the same time, large-scale and multilingual contexts show mixed or small effects of FL learning on maths achievement, often hinging on language proficiency and language-of-instruction alignment (González-Martín et al., 2024; Greisen et al., 2021). Recent reviews note that multilingual learners can perform below expectations when instruction or assessment occurs in a nondominant language, even though they may show enhanced mathematical encoding and processes in nonlanguage-related tasks (Dentella et al., 2024).

Despite these findings, research on first-time, school-based FL learning and mathematics achievement in adolescents remains scarce, particularly outside North America. This study addresses this gap by using international data to examine whether participation in FL programmes is associated with mathematics outcomes for 15-year-olds.

1.2. Foundations for effective foreign language learning

Effective FL learning depends on multiple factors beyond delivery method, including curriculum design, teaching strategies, educator expertise, and learner motivation (Dixon et al., 2012; Ryshina-Pankova, 2015). Among these, culturally enriched content is essential. Integrating cultural materials fosters communicative competence, critical thinking, and engagement with the target language (Lavrenteva & Orland-Barak, 2015). International frameworks (Council of Europe, 2001; National Standards, 2015) position culture as central to language learning, emphasising cross-cultural communication, intercultural awareness, and exposure to diverse perspectives as critical for proficiency (Chaika, 2023; Dlaska, 2000; Lusin et al., 2023).

Similarly, multicultural teaching practices – such as teaching sociocultural communication, addressing cross-cultural misunderstandings, and facilitating cultural exchanges – help students develop practical language skills (Byram, 2021; Byram et al., 2013; Cok, 2021; Savignon & Sysoyev, 2005). Finally, teacher expertise plays a pivotal role: experienced educators adapt lessons for diverse learners, foster inclusive classrooms, and implement evidence-based strategies that support language development and cognitive growth (Chan, 2006).

Ultimately, combining culturally enriched curricula, interactive methodologies, and skilled educators enhances language proficiency and intercultural competence, supporting cognitive skills that may translate into broader academic success.

1.3. Sociocultural and curricular moderators of mathematics achievement: A global perspective

In addition to cognitive skills, research has found several contextual factors that influence mathematical achievement during high school years. Socioeconomic status (SES), for instance, is a well-established determinant, often linked to lower academic performance among disadvantaged students across diverse educational systems and national contexts (Eriksson et al., 2021; Kim et al., 2019; OECD, 2023b).

Gender differences in mathematics achievement have also been widely studied. Several studies in different contexts have shown that when disparities are reported, their extent and nature vary significantly depending on cultural beliefs, educational policies, and societal norms (Hamamura, 2011; Hyde & Mertz, 2009; Nollenberger et al., 2016; Stoet & Geary, 2018).

Parental education also influences mathematical achievement, though studies differ on whether maternal, paternal, or combined parental education has the strongest impact (Chiu et al., 2016; Crede et al., 2015; Pishghadam & Zabihi, 2011). Similarly, parental involvement, particularly when involving high expectations and effective communication with schools, is often associated with positive outcomes (Boonk et al., 2018; Fiskerstrand, 2022; Hong et al., 2010). However, it has been noted that excessive supervision can yield adverse effects (Boonk et al., 2018; Hong et al., 2010; Rodríguez et al., 2017; Silinskas & Kikas, 2019).

Studies on school characteristics, including size and student–teacher ratios, present mixed evidence. Smaller classes and lower ratios are generally associated with improved performance, although other factors, such as school access to teaching resources, play a decisive role (Abizada & Seyidova, 2024; Olson et al., 2011). Additionally, urban students often outperform their rural peers, while students in independent schools tend to achieve higher scores than those in public schools in countries with dual education systems (Forgasz & Hill, 2013; Mohammadpour & Abdul Ghafar, 2014).

Beyond demographic and school-level factors, language policy and curriculum design add complexity in multilingual contexts. Mathematics outcomes vary with policy choices (e.g., language of

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

instruction), curriculum language demands, and student-body composition. For example, Luxembourg's European Public Schools – a system that allows families to choose the main language of instruction, typically to match their home language – report higher mathematics scores relative to schools following the traditional curriculum. This finding supports the idea that the alignment between home and school language and reduced language burden can enhance maths learning (Colling et al., 2024). Similarly, long-standing Canadian immersion studies show that teaching maths in a second language can yield comparable performance when proficiency is adequate and programmes are well-structured (Xu et al., 2022). These examples illustrate that mathematics outcomes are highly contingent on language match and programme design.

Understanding the impact of these contextual, demographic factors is essential for addressing disparities in maths performance and improving the accuracy of educational research. This is why large-scale international assessments offer a valuable platform for global comparisons, important for refining educational strategies and addressing inequalities.

1.4. Mathematical achievement and the Programme for International Student Assessment (PISA)

Mathematical proficiency is vital for educational attainment, career prospects, and individual and national development (Anderton et al., 2017; Delaney & Devereux, 2020; European Commission, 2011; Joensen & Nielsen, 2009). Better maths performance during high school has been correlated with improved transitions to higher education, stronger university performance, and enhanced career prospects (Anderton et al., 2017; Delaney & Devereux, 2020; European Commission, 2011; Joensen & Nielsen, 2009). Achievement in science, technology, engineering, and mathematics (STEM) fields, in particular, has been shown to drive innovation, economic growth, and societal welfare (Freeman et al., 2019). Despite its importance, maths outcomes have declined globally, with PISA reporting more countries experiencing drops than gains in 2018 – a trend that worsened in 2022, largely due to the COVID-19 pandemic (OECD, 2019b, 2023b).

PISA, a triennial assessment of 15-year-olds, measures reading, mathematics, and science performance, alongside career-related domains, such as financial literacy, which vary across iterations (OECD, 2019a). Since its launch in 2000, PISA has grown to include 80 countries and economies (OECD, 2024). PISA also gathers extensive contextual data from students, parents, teachers, and schools. Its rigorous design – standardised questionnaires, stratified sampling, advanced validation, and use of advanced statistical techniques to manage missing data – ensures comparability across diverse education systems (OECD, 2009, 2019a, 2024).

Given the cognitive, educational, and sociocultural benefits of FL learning, and the global concern over declining maths achievement, this study leverages PISA data to explore whether participation in FL programmes is associated with differences in mathematics performance among adolescents.

2. The current study

This study explores the relationship between school-based FL learning and mathematical achievement among 15-year-old students. It addresses two key gaps in the literature: the underrepresentation of adolescents – a critical developmental stage – and the predominance of North American research, which limits broader applicability.

Grounded in cognitive theory and empirical findings, this study models the association between FL learning and mathematics as indirect. It proposes, without empirically examining these mechanisms in the analyses, that FL learning may support mathematics achievement by strengthening EFs such as working memory, cognitive flexibility, and inhibitory control. It also considers a broader secondary theoretical contribution through higher-order thinking skills, including problem-solving and abstract reasoning. These mechanisms are presented as conceptual explanations rather than tested mediators.

Recognising the influence of sociocultural and curricular factors, our study focuses on students enrolled in mainstream programmes where FL is taught as a subject, excluding immersion and bilingual contexts to avoid confounding variables. For analysis, we operationalised "language learners" as students whose reported home language matched their language of instruction.

The research addresses two primary questions:

(1) To what extent does the duration of school-based FL learning is linked to the mathematical achievement of 15-year-old students, after controlling for established determinants of academic performance?
(2) To what extent do specific characteristics of FL learning programmes predict mathematical achievement in this population?

3. Methodological design

3.1. Data selection

This study utilised data from the Organisation for Economic Co-operation and Development (OECD) PISA 2018 cycle to address the research questions. PISA assesses the academic knowledge and skills of 15-year-old students across various domains, including reading, mathematics, and science, while also collecting detailed background information through student, parent, teacher, and school-level questionnaires.

The 2018 dataset, which includes data from over 600,000 students across 79 countries and economies, is particularly valuable as it represents the first and, at the time of writing, only instance where intercultural communication skills, including FL learning, were included in the assessment (OECD, 2019a). This allowed for an in-depth analysis of these skills alongside student performance in mathematics.

To ensure alignment with the research questions and maintain data validity, a subsample of students was selected based on language-use criteria and data completeness. Our study focuses on language learners in mainstream programmes where FLs are taught as subjects, rather than through immersion. To minimise the likelihood of including proficient bilinguals – who, according to prior research, may exhibit different cognitive profiles (White & Greenfield, 2017) – only students who reported using the same language at home and at school were included in the sample. For example, if a student's home language (e.g., Spanish) differed from the language of assessment (e.g., English), their data were excluded. Additionally, cases with missing or invalid responses for key variables, as well as extreme outliers, were removed to meet model requirements and reduce bias in the results (see Supplementary Material Appendices S1 and S2 for details).

For the first research question – examining the relationship between time spent on school-based FL learning programmes and mathematics performance – we used data from the student, parent, and school questionnaires. After applying our filters, 34% of cases in the PISA dataset were removed due to invalid or incomplete data, 13% were excluded based on the language-mismatch filter, and a further 4% were excluded due to the outlier removal process.

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

Consequently, the final sample for the first research question comprised 300,656 students from 73 countries and economies.

The second research question focused on the role of specific FL programme characteristics in predicting mathematics performance. This analysis incorporated all datasets in the first study, supplemented with data from teacher questionnaires. To ensure conceptual relevance, the dataset was filtered to include only language teachers, as the teacher-level variables incorporated in the models – such as years of experience, multicultural teaching practices, and cultural competence – are specific to FL instruction and would not apply to teachers of other subjects. Because only 19 countries administered the teacher questionnaire and the proportion of language teachers was relatively small, the final sample for the second question was substantially smaller, consisting of 53,459 students and 16,342 teachers, representing 31% and 23% of their respective total samples. Further details on the data selection and preparation processes are provided in the Supplementary Material (Appendices S1 and S2).

3.2. Variable description

3.2.1. Foreign language learning variables

Our primary predictor variable was FL learning time in minutes per week (*FLMINS*). *FLMINS* serves as a proxy for exposure to FL learning, as direct measures of proficiency were unavailable.

This variable was calculated by combining two PISA 2018 variables: FL periods per week (*ST059Q04HA*) and average minutes per class period (*ST061Q01NA*), both derived from students' input in the student/parent questionnaire.

The calculation was performed as follows:

$$FLMINS = (ST059Q04HA) \times (ST061Q01NA).$$

To ensure reliability, outliers were addressed using two approaches. First, *FLMINS* values were capped at the total minutes of instruction (*TMINS*), an OECD-derived index reflecting realistic instructional time. Second, within each school, the interquartile range (IQR) method was applied to exclude extreme values exceeding the upper bound (James et al., 2023). This ensured that only *FLMINS* values disproportionately high relative to other values within the same school were removed. Tables 1 and 2 provide a summary of all the variables used, including their description, sources, and values.

3.2.2. Mathematical performance variables

PISA assesses students' mathematical achievement using a set of 10 plausible values (PVs) rather than a single score. These PVs are not raw scores but random draws from a student's estimated ability distribution, derived from latent proficiency models that incorporate extensive contextual covariates (Jewsbury et al., 2024). This approach addresses the fact that, to encourage participation and minimise fatigue, each student completes only a subset of the total pool of mathematics questions during a given test cycle (OECD, 2009).

The PVs reflect students' abilities across various mathematical domains, including formulation, application of concepts, facts, and procedures, as well as reasoning and interpretation of mathematical results (OECD, 2019a). In this study, the 10 PVs serve as the primary outcome variables. Following OECD guidance, all PVs were included in the analyses to account for their statistical nature and to ensure validity and robustness (OECD, 2009).

To enable comparability across cycles and countries, PISA standardises students' PVs to have a mean of 500 and a standard deviation (SD) of 100 points (OECD, 2009). In 2018, due to a decline in student performance, the OECD average in mathematics was 489, with an SD of 91 points (OECD, 2019b). This process ensures that the data are also suitable for international comparative analysis.

3.2.3. School-related variables

Type of Admission (ADMISSION). It captures potential differences in student achievement stemming from selective school entry. Based on school responses to PISA variable *SC012Q01TA*, it reflects the frequency of admission based on academic performance (including placement tests). Its inclusion helps control bias from academically elite cohorts that might otherwise confound the relationship between our variables of interest.

Mathematics learning time (MMINS). As maths instructional time (in minutes per week) is logically associated with mathematical performance, *MMINS* was included as a control variable to account for its potential confounding effect on the relationship between FL learning time (*FLMINS*) and mathematical achievement.

MMINS, an OECD index, is based on student and school responses regarding the number of mathematics periods per week and the average length of each period (OECD, 2020).

School location (SCHLOC). Reported by the OECD using data from the school questionnaire, *SCHLOC* categorises schools based on population size, from villages to large cities, to control for geographical disparities.

School type (SCHTYPE). This variable is recorded by the OECD, based on school records, and groups them as private-independent, private government-dependent, or public. *SCHTYPE* was incorporated into the model to account for differences across school ownership.

Student–teacher ratio (STRATIO). Drawn from the school dataset, *STRATIO* is an OECD index constructed from principals' reports regarding average class size It serves as a proxy for class size, a well-established determinant of maths performance.

School size (SCHSIZE). Also sourced from the school dataset, *SCHSIZE* is a school-level OECD index representing the total number of enrolled students. This variable was included to capture the potential effect of school scale on student achievement.

Teacher experience in years (TEACHEXP). Derived from the teachers' dataset, *TEACHEXP* represents the total years of teaching experience, accounting for variations in instructional quality across classrooms.

3.2.4. Teaching programme variables

The following variables were included to ensure that the effects of FL learning are assessed independently of the broader benefits of engagement with multicultural teaching and programmes.

Multicultural curriculum (MCCUR). This index was derived from five PISA school questionnaire items (*SC167Q01HA–SC167Q05HA*) asking whether the formal curriculum for 15-year-olds includes components such as intercultural communication, cultural knowledge, openness to intercultural experiences, and FL instruction. Each item was coded 1 = Yes, 0 = No, and equally weighted. The score represents the proportion of these elements present at the school level, yielding a range from 0 to 100:

$$\text{MCCUR} = \frac{\text{Number of "Yes" responses}}{5} \times 100.$$

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

Table 1. Description of variables: linear mixed model variables (Study 1)

Variables	PISA 2018 items	Description	Values
Main predictor			
FLMINS	Composite* (*ST059Q04HA, ST061Q01NA*)	Foreign language learning time in minutes a week	Time in minutes calculated using – number of class periods per week in foreign language (*ST059Q04HA)* – average minutes per class period (*ST061Q01NA)*
Outcome			
pv_math1–10	*PV1MATH* to *PV10MATH*	Plausible values in mathematics	Standardised with an average score of 489, with an SD of 91 points
Identifiers			
CNT	*CNT*	Country code 3-character	3-character code (e.g., ESP for Spain, AUS for Australia)
CNTSCHID	*CNTSCHID*	International school identifier	6- to 8-digit code (2- to 4-digit country code + 4-digit individual identifier)
CNTSTUID	*CNTSTUID*	International student identifier	
Filters			
LANGTEST	*LANGTEST_COG*	Language of assessment	3-digit code (e.g., 156 for Spanish, 313 for English)
LANGN	*LANGN*	Language at home	
TMINS	*TMINS*	Total learning time	Time in minutes (OECD index)
School			
ADMISSION	*SC012Q01TA*	Student admission to school based on record of academic performance (including placement tests)	1 = never 2 = sometimes 3 = always
MMINS	*MMINS*	Mathematics learning time in minutes a week	Time in minutes (OECD index)
SCHLOC	*SC001Q01TA*	School location, based on population	1 = village or rural area (< 3,000 people) 2 = small town (3,000 to 15,000 people) 3 = town (15,000 to 100,000 people) 4 = city (100,000 to 1,000,000 people) 5 = large city (> 1,000,000 people)
SCHLTYPE	*SCHLTYPE*	School type, based on school ownership	1 = private independent 2 = private government-dependent 3 = public
Student			
ESCS	*ESCS*	Economic, Social and Cultural Status	Above and below 0 (OECD average) with an SD of 1 (OECD index)
GENDER	*ST004D01T*	Student gender (standardised)	1 = female 2 = male
GRADE	*ST001D01T*	Student standardised international grade	7 = grade 7 or equivalent 8 = grade 8 or equivalent 9 = grade 9 or equivalent 10 = grade 10 or equivalent 11 = grade 11 or equivalent 12 = grade 12 or equivalent 13 = grade 13 or equivalent
Parents			
MISCED	*MISCED*	Maternal education level	0 = none 1 = ISCED 1 (primary education) 2 = ISCED 2 (lower secondary education) 3 = ISCED 3B and 3C (secondary education for workforce) 4 = ISCED 3A and 4 (postsecondary non-tertiary education) 5 = ISCED 5B (short-cycle skill-based tertiary education) 6 = ISCED 5A and 6 (bachelor's degree or equivalent level)
FISCED	*FISCED*	Paternal education level	

*Composite variables are calculated as described in the Methods section and Supplementary Material.

This approach follows PISA's aggregation conventions (OECD, 2020) and reflects the extent to which multicultural content is formally integrated into the curriculum. Moreover, all components are derived from question SC167 and designed to be interpreted together, ensuring consistency in meaning. The inclusion of these elements aligns with research emphasising the role of cultural education in supporting language learning and cognitive development (Dlaska, 2000).

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

Table 2. Description of variables: random forest variables (Study 2)

Variables	PISA 2018 items	Description	Values
Identifiers			
CNTTCHID	*CNTTCHID*	International teacher identifier	6- to 8-digit code (2- to 4-digit country code + 4-digit individual identifier)
Filters			
TEACHERID	*TEACHERID*	Type of teacher identification code	4 = reading/language specific 5 = general
School			
STRATIO	*STRATIO*	Student–teacher ratio, average per school	1 to 100 (OECD index)
SCHSIZE	*SCHSIZE*	School size (sum) in number of students	Total number of students (OECD index)
Teaching			
TEACHEXP	*TC007Q02NA*	Teaching experience in years	Total teaching years of experience
GCSELF	*GCSELF*	Teacher's self-efficacy in multicultural environments	Above and below 0 (OECD average) (OECD index)
MCCUR	Composite* (*SC167Q01HA–SC167Q05HA*)	Multicultural curriculum	Percentage of "Yes" responses (0–100%) indicating formal curriculum for: – communicating with people from different cultures (*SC167Q01HA)* – knowledge of different cultures (*SC167Q02HA)* – openness to intercultural experiences (*SC167Q03HA)* – respect for cultural diversity (*SC167Q04HA)* – foreign languages (*SC167Q05HA)*
MCTEACH	Composite* (*SC159Q01HA, SC165Q02HA, SC165Q06HA–SC165Q09HA, TC207Q05HA*)	Multicultural teaching practices	Percentage of "Yes" responses (0–100%) indicating implementation of these practices: – hosting teachers from other countries (*SC159Q01HA)* – teaching the history of international cultural groups (*SC165Q02HA)* – offering exchange programme with schools abroad (*SC165Q06HA)* – organising multicultural events (*SC165Q07HA)* – celebrating festivities from other cultures (*SC165Q08HA)* – encouraging students to communicate with people from other cultures (*SC165Q09HA)* – promoting foreign language skills (*TC207Q05HA*)
Parents			
PARINV	Composite* (*SC064Q01TA, SC064Q02TA*)	Parental involvement in students' progress	Percentage (0 to 100%) indicating the proportion of parents who – discussed their child's progress with a teacher on their own initiative (*SC064Q01TA)* – discussed their child's progress on the initiative of one of their child's teachers (*SC064Q02TA)*

Note: All variables from Study 1 are also included in Study 2.
*Composite variables are calculated as described in the Methods section and Supplementary Material.

Multicultural teaching practices (MCTEACH). This composite index captures the extent to which schools implement practices that actively promote cross-cultural interaction and understanding, such as hosting teachers from other countries, teaching the history of international cultural groups, offering exchange programmes, celebrating cultural festivities, encouraging intercultural communication, and promoting FL skills (Cok, 2021; Savignon & Sysoyev, 2005).

The index was calculated from seven PISA items (*SC159Q01HA*, *SC165Q02HA*, *SC165Q06HA–SC165Q09HA*, *TC207Q05HA*). Each item was coded 1 = Yes, 0 = No, and equally weighted. The composite score represents the proportion of these practices implemented at the school level, yielding a range from 0 to 100:

$$\text{MCTEACH} = \frac{\text{Number of "Yes" responses}}{7} \times 100.$$

Internal consistency was high (Cronbach's $\alpha = 0.923$), supporting the reliability of this measure.

Teacher's self-efficacy in multicultural environments (GCSELF). It is an OECD-calculated index reflecting teachers' confidence in addressing multicultural challenges, based on self-assessment

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

across various dimensions such as raising awareness of cultural differences and managing multicultural classrooms effectively.

3.2.5. Student-related variables

Socioeconomic status (ESCS). To ensure the analysis examines the effects of *FLMINS* independently of socioeconomic disparities, the Economic, Social and Cultural Status *(ESCS)* index was included in the model. This index is a composite measure developed by the OECD, based on answers from both student and parent questionnaires. It reflects students' family backgrounds by combining parents' highest education levels, highest occupational status, and home possessions.

Notably, the "cultural" aspect of the *ESCS* index pertains to intellectual resources and academic achievement potential, rather than social customs or ethnic-specific behaviours. This component reflects the educational background and cognitive environment within the family. For example, home possessions such as books or access to technology are included as proxies for intellectual capital, which supports cognitive development and learning opportunities (OECD, 2020).

Gender (GENDER). Recorded directly from the student dataset, it was derived from students' self-reports. This categorical variable was included to control for gender-related differences in mathematics performance.

Academic year level (GRADE). PISA assesses students aged 15 years and 3 months to 16 years and 2 months (OECD, 2020), which often corresponds to one or two years before high school graduation. However, year levels vary significantly across countries, which may introduce differences in students' maturity and preparation for assessments. Including *GRADE* in the analyses ensures comparability of participants across educational systems.

3.2.6. Parent-related variables

Parental level of education (MISCED and FISCED). Maternal *(MISCED)* and paternal *(FISCED)* education levels are reported by PISA using ISCED classifications (UNESCO, 2021). These variables were included to add nuance to the family socioeconomic background.

Parental involvement (PARINV). This continuous variable was constructed for this study using two PISA items: the percentage of parent–school discussions initiated by parents (*SC064Q01TA*) and those initiated by teachers (*SC064Q02TA*). To capture the highest level of engagement, *PARINV* represents the greater of the two percentages. This measure reflects the extent of parental communication with schools, which is often associated with improved academic outcomes (Boonk et al., 2018).

4. Study 1. Effect of *FLMINS* on maths scores

4.1. Method: Linear mixed modeling

To investigate the association between FL learning time (*FLMINS*) and students' mathematics performance, a linear mixed-effects model (LMM) was fitted using the *lme4* package in R (Bates et al., 2015). This model accounts for the nested structure of the data, where students are grouped within countries, and it controls for sources of variability at both levels.

The model was specified as follows:

$$\begin{aligned} \textit{Math Score} \sim\ & \textit{FLMINS} + \textit{MMINS} + \textit{GENDER} + \textit{GRADE} \\ & + \textit{ESCS} + \textit{MISCED} + \textit{FISCED} + \textit{SCHLOC} \\ & + \textit{SCHLTYPE} + \textit{ADMISSION} \\ & + (1|\text{CNTSTUID}) + (1|\text{CNT}). \end{aligned}$$

This specification estimates the association between *FLMINS* and mathematical performance (*Math Score*), while controlling for time spent on mathematics (*MMINS*), demographic factors (gender, grade), socioeconomic and cultural background (*ESCS*), parental education level (*MISCED*, *FISCED*), and school characteristics (location, type, and admission process). Random intercepts at the student (*CNTSTUID*) and country (*CNT*) levels account for individual differences and country-level variability, respectively. This structure addresses unobserved heterogeneity within countries and among individual students.

An initial screening of *FLMINS* revealed the presence of outliers, even after applying data cleaning filters described earlier. To address this, *FLMINS* was winsorised at 1,500 minutes per week, the maximum value for mathematics instruction time (*MMINS*). This threshold aligns with educational practice, where FL instruction typically receives fewer periods compared to core subjects like mathematics (OECD, 2023a). Winsorisation, a statistical technique that limits extreme values in a dataset by replacing outliers with a specified value, was employed to preserve meaningful data points – such as those from schools prioritising FL study – while preventing *FLMINS* to exceed realistic bounds and having disproportionate influence on analyses (Gelman & Hill, 2006; Tabachnick & Fidell, 2018; Wilcox, 2017).

In addition, we tested for a potential nonlinear relationship between *FLMINS* and *Math Score* using a generalised additive model (GAM) with the *mgcv* package for R (Wood, 2017). The analysis indicated a linear relationship between the variables of interest, supporting the use of an LMM (for details, see Supplementary Material Appendix S2). Given the size and complexity of the data, the use of a GAM, instead of an LMM, would have introduced substantial computational challenges without significantly improving model performance.

Importantly, school size (*SCHSIZE*) and class size (*CLSIZE*) were initially considered as predictors but were ultimately excluded, as their inclusion would have removed all countries for which the OECD did not report or compute values for these variables. Because complete data were required to estimate the LMM, including *SCHSIZE* and *CLSIZE* would have substantially reduced the sample without offering commensurate analytical benefit. To include the effect of class numbers without significantly reducing the sample, student–teacher ratio (*STRATIO*) was used instead. Other variables were initially considered but later removed, such as student age, which was replaced by *GRADE*, and highest parental income, which was replaced by parental education level (*MISCED, FISCED*).

4.2. Study 1 results

Given the potential associations among predictors, Pearson's correlations (Olivoto & Lúcio, 2020) and variance inflation factors (VIFs) (Lüdecke et al., 2021) were calculated to assess multicollinearity. Although some variables exhibited moderate-to-strong correlations ($r = .70$ for *ESCS*/*MISCED*, $r = .67$ for *ESCS*/*FISCED*), no significant collinearity concerns were identified (VIFs: 2.43 for *ESCS*, 1.78 for *FISCED*, and 1.89 for *MISCED*) (for details, see Supplementary Material Appendix S2).

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

Table 3. Linear mixed-effects model results: fixed effects estimates and random effects intercepts

Predictors	Estimates	CI	*p*
Fixed effects			
(Intercept)	173.17	161.05–185.29	<.001
Foreign language learning time (*FLMINS w*)	0.08	0.07–0.08	<.001
Mathematics learning time (*MMINS*)	0.03	0.03–0.03	<.001
Gender (*GENDER*)	10.43	9.92–10.93	<.001
Academic year level (*GRADE*)	28.08	27.57–28.60	<.001
Socioeconomic level index (*ESCS*)	29.33	28.91–29.76	<.001
Maternal level of education (*MISCED*)	−2.16	−2.40 to −1.93	<.001
Paternal level of education (*FISCED*)	−0.83	−1.05 to −0.60	<.001
School location (*SCHLOC*)	6.31	6.07–6.55	<.001
School type (*SCHLTYPE*)	−5.20	−5.68 to −4.73	<.001
Type of admission (*ADMISSION*)	2.42	2.18–2.66	<.001
Random effects			
σ2	1,349.73		
τ00 CNTSTUID	4,825.45		
τ00 CNT	2,192.65		
ICC	0.84		
N CNTSTUID (students)	300,656		
N CNT (countries)	73		
Observations	3,006,560		
Marginal R^2/conditional R^2	0.162/0.865		

The linear model, which included both fixed and random effects, achieved a conditional R^2 of .865, while the marginal R^2 was .162. This indicates that the inclusion of random intercepts for students and countries considerably improved the model fit, with 86.5% of the variance in mathematics scores explained by the full model compared to 16.2% achieved by fixed effects alone. Table 3 presents a summary of the results obtained.

All fixed effects were statistically significant ($p < .001$), which was expected given the large sample size, suggesting a robust relationship between the predictors and mathematics scores. *FLMINS* showed a positive association with maths scores, estimated at 4.8 PISA points higher per additional hour of FL instruction (or $b = .08 \times 60$ minutes). For context, across the middle 50% (110 to 240 min/week), this effect corresponds to a positive variation of 8.8–19.2 points in predicted achievement.

Note that some of the variables included in the analysis were categorical – specifically *GENDER*, *GRADE*, *MISCED*, *FISCED*, *SCHLOC*, *SCHLTYPE*, and *ADMISSION* – and were treated accordingly. However, instead of presenting the individual effects for each category level, we only report their combined effect.

Among the demographic variables, SES (*ESCS*) and *GRADE* had the strongest effects. A one-standard-deviation increase in *ESCS* corresponded to a 29.33-point difference in mathematical scores – for the middle 50% of the sample (−0.86 to +0.66), this represents a 44.5-point spread. Similarly, each additional grade level was associated with an estimated difference of 28.08 points (e.g., 56.16 points between grades 8 and 10).

Other school-level factors showed smaller but significant associations. Schools located in larger cities (*SCHLOC* = 5) reported scores 25.24 points higher than those in rural areas (*SCHLOC* = 1), averaging a 6.31-point difference in scores per level. Likewise, students attending private independent schools (*SCHLTYPE* = 1) showed scores 10.40 points higher than those attending public schools (*SCHLTYPE* = 3) and about 5.20 points higher than those in government-dependent private schools. Similarly, academic selectivity (*ADMISSION*) accounted for a 4.84-point difference between schools that "always" versus "never" use placement tests ($b = 2.42$).

The remaining variables showed lesser effects on mathematics scores. Results indicate that the time spent learning mathematics (*MMINS*) had a limited positive association, corresponding to about 1.8 points per hour, a significantly smaller effect than the time spent learning FL. Gender differences were similarly modest, with boys scoring 10.43 points higher than girls on average. Surprisingly, higher maternal education levels (*MISCED*) was associated with an average difference of −2.16 point, while each increase in paternal education level (*FISCED*) corresponded to an estimated difference of −0.83 point in mathematics scores.

To explore differences among countries, these were grouped by economic status according to the latest United Nations *World Economic Situation and Prospects* report (United Nations, 2025). This approach was selected since national economic growth has been shown to mediate the effect of SES – the strongest predictor in our main model – on academic achievement (Kim et al., 2019). Research also indicates there is a reciprocal relationship between economic growth and academic performance, further emphasising the rationale for this stratification (Hanushek & Woessmann, 2021; Valero, 2021; Woessmann, 2016).

This model revealed a consistent positive association between increased FL instruction time and mathematics performance across all economic categories (Figure 1). However, the magnitude of this effect varied notably by economic status. Economies in transition, countries such as Azerbaijan, Belarus, and Georgia, exhibited the steepest slope, with mathematics scores on average 11.4 points higher per additional hour of FL learning per week. In contrast, developing countries (e.g., Argentina, Morocco, and Vietnam) recorded the lowest baseline scores and the smallest overall estimated difference, with an average of 4.2 points per hour of FL instruction. Developed countries (e.g., Canada, Japan, and Norway) also displayed a positive trend, with an average difference of 7.2 points per hour of FL learning (for details, see Supplementary Material Appendix S2).

4.3. Robustness tests

To assess the stability of the observed association between foreign-language learning and mathematics achievement, we conducted additional analyses using alternative subsamples and model specifications.

4.3.1. Multilingual subsample

For this check, the model was re-estimated using only data from multilingual education systems (Luxembourg, Switzerland, Canada, Finland, Spain, and Singapore), keeping all other filters in place. The coefficient for FL instruction time remained positive ($b = 0.05$, $p < .001$), corresponding to approximately 3 PISA points per additional weekly hour of FL learning. Although statistically

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

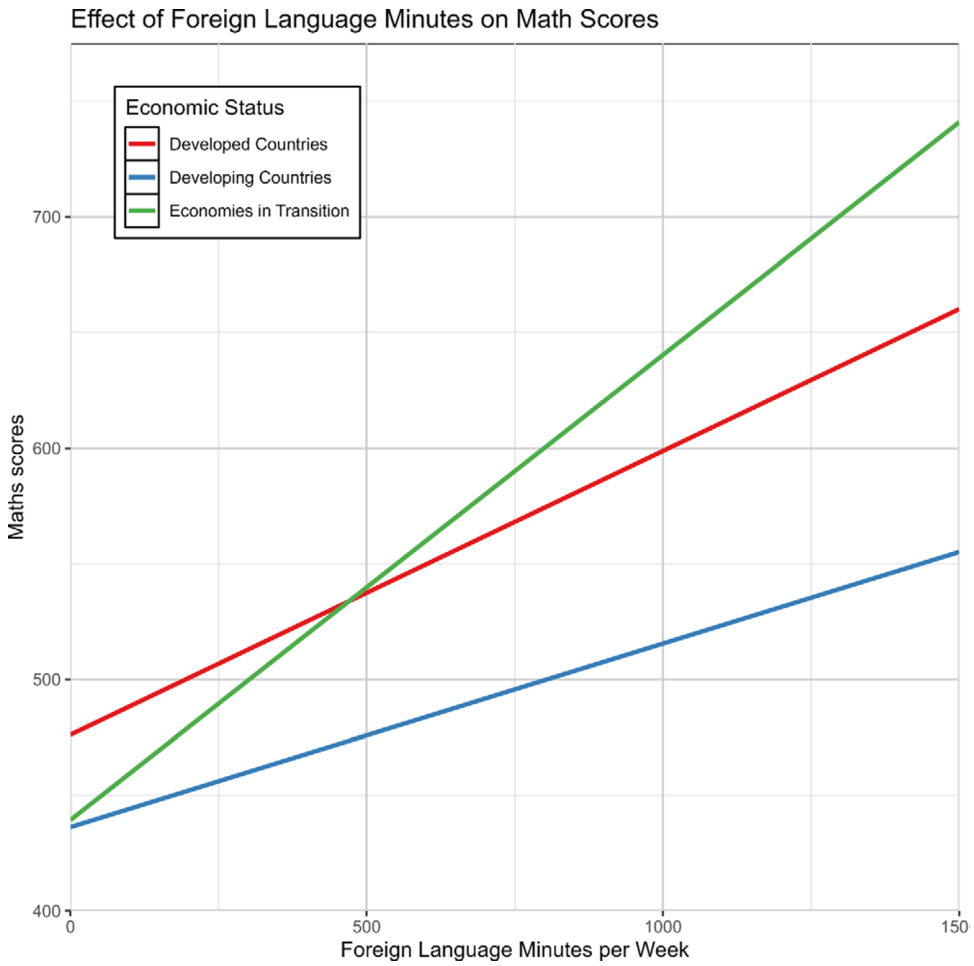


Figure 1. Linear mixed effects model results: Effect of FLMINS on mathematics scores according to countries' economic status.

reliable, this effect was modest compared to SES and grade level, indicating that contextual factors exert a stronger influence.

4.3.2. Expanded sample with language-mismatch indicator.

This test was performed by removing the language-use filters and adding a binary indicator for language mismatch (1 if home = test language, 0 if not), while keeping the covariates identical to the main specification. This expanded sample (382,647 students, 95% of PISA 2018 complete cases) yielded consistent results: The association between *FLMINS* and mathematics performance remained positive and significant ($b = 0.07$, $p < .001$; 95% confidence interval (CI) [0.06, 0.07]), corresponding to approximately 4.2 PISA points per additional hour of FL instruction per week. The language mismatch indicator was a significant predictor ($b = 5.7$, $p < .001$), confirming its influence on achievement but not accounting for the *FLMINS* effect.

4.3.3. Within- vs between-school decomposition (Mundlak approach)

The Mundlak approach (Mundlak, 1978) was applied to assess if the observed FL–maths association persisted at the student level or was a reflection of school-level differences (e.g., wealthier schools offering more language classes). This method separates within-school (comparing students to their classmates) and between-school (comparing school averages) effects, ensuring that variations in school-level allocation do not inflate the individual-level estimate.

The within-school coefficient remained positive and statistically significant ($b = 0.06$), indicating that a student receiving one additional weekly hour of FL instruction (relative to their school peers) is predicted to score 3.6 points higher in mathematics. The between-school component was also positive ($b = 0.09$), showing that students attending schools that offer an additional weekly hour of FL instruction tend to score 5.5 points higher in maths tests. While these effects are smaller than those of SES and grade level, the persistence of the within-school association indicates that the relationship is not solely a byproduct of school placement.

Full results for all robustness analyses and alternative specifications are provided in the Supplementary Material (Table S3 and Appendix S2).

4.4. Study 1 discussion

Study 1 examined the relationship between foreign language instruction time (*FLMINS*) and mathematics achievement using a linear mixed model (LMM). The main model revealed a positive association: Mathematics scores were approximately 4.8 points higher per additional weekly hour of foreign-language (FL) instruction. This finding reinforces prior research on the

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

academic benefits of FL learning (Nucette et al., 2024; Woll & Wei, 2019), as well as previous research on the indirect link between FL learning and improved numeracy performance, mediated by EF (Al-khresheh & Karmi, 2024; Shoghi & Ghonsooly, 2018).

Although the model explained a significant portion of the variance in mathematical scores (R^2 = 86.5%), the analysis highlighted that most of the variability is attributable to country and individual differences, particularly socioeconomic level and grade. School-level characteristics (e.g., school location, type), while influential, played a comparatively smaller role in the variation.

Robustness checks further strengthened these findings against methodological challenges. The association persisted, though at a reduced magnitude (3 PISA points per FL-hour/week), when the analysis was restricted to multilingual education systems, indicating the effect is not confined to language learning contexts. Expanding the sample to include language-mismatch cases confirmed the overall stability of the association and indicated that it is not driven by language-background selection or test-language disadvantage. Crucially, decomposing exposure into within- and between-school components (Mundlak adjustment) revealed that the effect holds true among peers in the same school (b = 0.06) and across schools based on their *FLMINS* allocation (b = 0.09). These patterns suggest that the observed relationship is not only attributable to fixed differences between schools or non-random placement of students. While they do not establish causality, they are consistent with our theoretical interpretation that FL learning may support mathematics achievement indirectly through cognitive processes.

Stratification by country-level economic status demonstrated that *FLMINS* was consistently associated with higher mathematical performance across all economic categories. The strongest effects were observed for economies in transition, an outcome that could be attributable to specific pedagogical mechanisms, similar educational reforms, or shared attitudes toward FL learning. Another possible explanation is cultural in nature, with intrinsic motivation and resilience playing a role (Huisman et al., 2018). As neither of these mechanisms was the primary focus of the present study, additional research is needed to explore these hypotheses. In contrast, developing and developed economies exhibited smaller gains. In developing economies, this may reflect structural constraints such as limited access to quality education or inadequate teacher training, whereas in developed economies, well-established teaching practices and curricula may lead to smaller effects. These findings align with previous literature highlighting the influence of national economic contexts on educational outcomes (Hanushek & Woessmann, 2017).

While the positive association between *FLMINS* and mathematics achievement was robust across specifications, it should be interpreted as correlational rather than causal, given the cross-sectional nature of the data. We acknowledge that unobserved school- and country-level factors – such as specific teaching practices, student–teacher rapport, curriculum design, and educational policies – likely contribute to the observed relationship and require further exploration.

To address the limitations of traditional approaches, a machine learning (ML) framework was adopted for the second research question. ML techniques offer enhanced capacity to manage complex datasets, tolerate noise, and identify influential predictors with high accuracy – making them particularly suitable for analysing large-scale educational datasets such as PISA. This approach also enables cross-validation of findings from Study 1, contributing to the robustness and generalisability of the results.

5. Study 2. Effect of foreign language programme characteristics on maths scores

5.1. Method: Machine learning approach (random forest)

To evaluate the impact of FL programme characteristics on mathematical achievement, an ML approach was adopted, specifically the random forest (RF) method, implemented via the R package *randomForest* (Breiman, 2001; Liaw & Wiener, 2002).

The primary rationale for selecting this method was its high predictive accuracy. ML methods, particularly tree-based models, are designed to construct statistical models that generalise effectively to unseen data, whereas traditional models often perform optimally on the data to which they were fitted. In other words, ML models tend to produce predictions that closely align with actual outcomes (Hastie et al., 2009). Due to their predictive power, such algorithms have been widely applied across various fields, including psychology and education, to uncover patterns that might otherwise remain undetected (Balabied & Eid, 2023; Liew et al., 2025; Nachouki et al., 2023; Yarkoni & Westfall, 2017).

RF algorithms, along with other ML approaches, have been increasingly applied to identify predictors of mathematics and science literacy, using large-scale international student assessments, such as PISA and Trends in International Mathematics and Science Study (TIMSS) (Arroyo Resino et al., 2024; Gil-Madrona et al., 2025; Song & Cutumisu, 2025). Their growing use likely reflects their suitability for large and complex datasets. These algorithms also demonstrate a greater resilience to missing data and are less influenced by outliers, ensuring robust and reliable results (Hastie et al., 2009; Liaw & Wiener, 2002). RF models also provide an important interpretative advantage by quantifying the relative importance of predictor variables, offering insights into their practical significance on the outcome (James et al., 2023; Varian, 2014; Yarkoni & Westfall, 2017). These features make them superior to more traditional statistical models, including LMMs, which have known issues to detect complex interactions and may overestimate the significance of certain variables in large samples (Lin et al., 2013; Thiese et al., 2016).

Building on these strengths, the present study applies an RF model to examine a broader range of predictors than those considered in Study 1 by incorporating factors related to FL teaching programmes, such as multicultural curriculum and teaching practices. The primary objective was to identify the variables that most robustly and accurately predict the outcome variable – *Math Score.* Accordingly, the model included established demographic predictors of academic achievement, alongside variables specific to FL programmes (see Table 2).

The predictors can be categorised as follows:

- **Student-level predictors**: Gender, grade level, SES (*ESCS*), parental involvement (*PARINV*), and parents' highest educational levels (*MISCED, FISCED*).
- **Foreign language programme characteristics**: Time allocated to FL study (*FLMINS*), teacher experience in years (*TEACHEXP*), teacher self-efficacy in multicultural environments (*GCSELF*), and indicators of multicultural teaching practices and curriculum (*MCTEACH*, *MCCUR*).
- **School-level predictors**: Student–teacher ratio (*STRATIO*), school size (*SCHSIZE*), school ownership (*SCHTYPE*), school

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

location (*SCHLOC*), and mathematics instructional time (*MMINS*) as a control for time-on-task effects.

The dataset was randomly split into training (80%) and testing (20%) subsets. This approach ensured that model fit indices reflected the model's ability to predict outcomes on unseen data (Varian, 2014), thereby providing external validation of model performance (hold-out testing). Internal model performance was evaluated using out-of-bag (OOB) validation, which provides a cross-validation estimate based on bootstrap resampling during training. Performance was quantified using root-mean-square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2).

5.2. Study 2 results

The RF model performed well across both internal and external validation methods. The OOB cross-validation yielded an R^2 of 0.962, an RMSE of 17.29, and an MAE of 9.43. Performance on the independent hold-out test set was nearly identical (R^2 = 0.966, RMSE = 17.16, MAE = 9.41), indicating minimal overfitting and strong generalisability within the PISA dataset. Given the range of mathematics scores (136–799.6) and the mean of 485.1 in the data, the RMSE and MAE correspond to deviations of approximately 3.5% and 1.9% of the mean score, respectively.

These results indicate that the model explained approximately 96% of the variance in the OECD-generated PVs for mathematics scores, with a low prediction error, as illustrated in Figure 2.

Predictor importance was assessed using the percentage increase in mean-squared error (%IncMSE), a metric that quantifies the reduction in the model's predictive performance when the values of a given variable are permuted while keeping all other variables unchanged. A higher %IncMSE value indicates greater predictor importance of a particular variable, as altering its values results in an error of higher magnitude (Breiman, 2001; Liaw & Wiener, 2002).

Using this approach, Economic, Social, and Cultural Status (*ESCS*) emerged as the most influential predictor (%IncMSE = 364.9%). This means that randomly changing *ESCS* values, while keeping all other variables constant, introduces an error 3.6 times greater than that of the original model. The second most important predictor, closely following *ESCS*, was *FLMINS* (363.3%), further confirming its strong association with mathematics performance.

Other important predictors included the implementation of multicultural curricula (*MCCUR* = 277.9%) and maths instruction time (*MMINS* = 274.9%), while *SCHSIZE* (259.4%) and

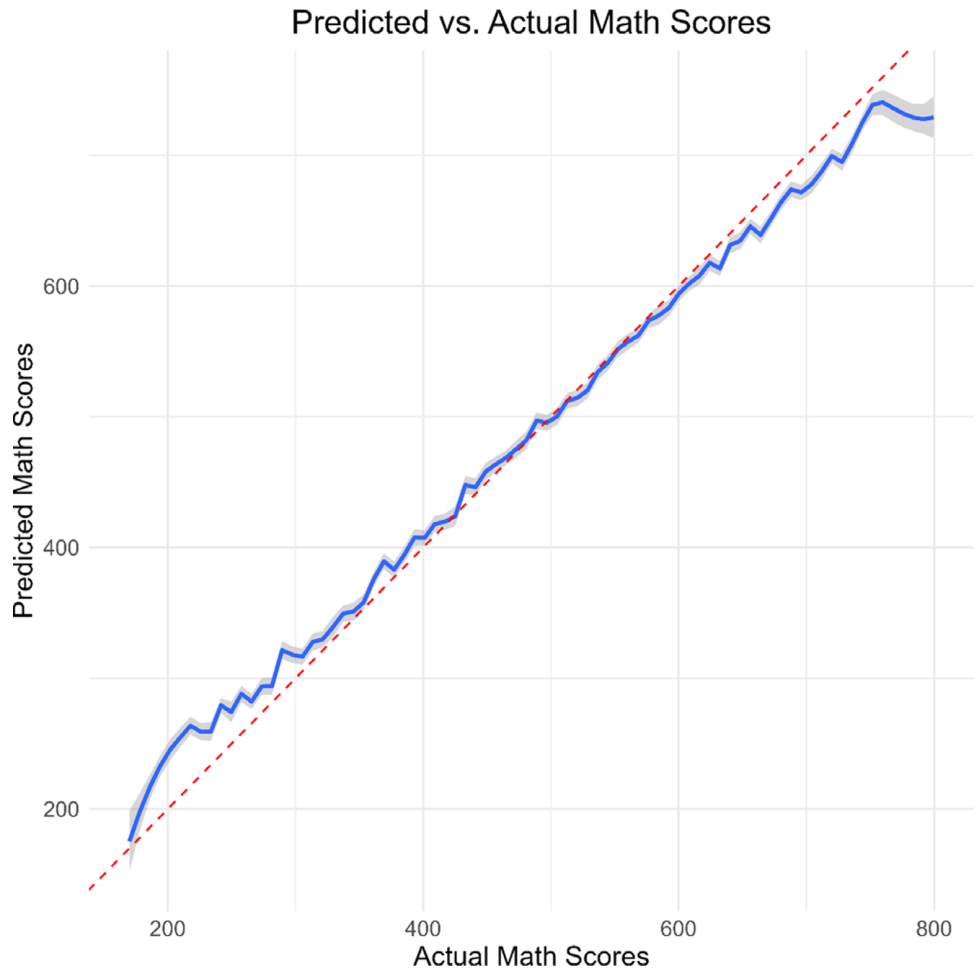


Figure 2. Random forest model results: Predicted vs actual values.

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

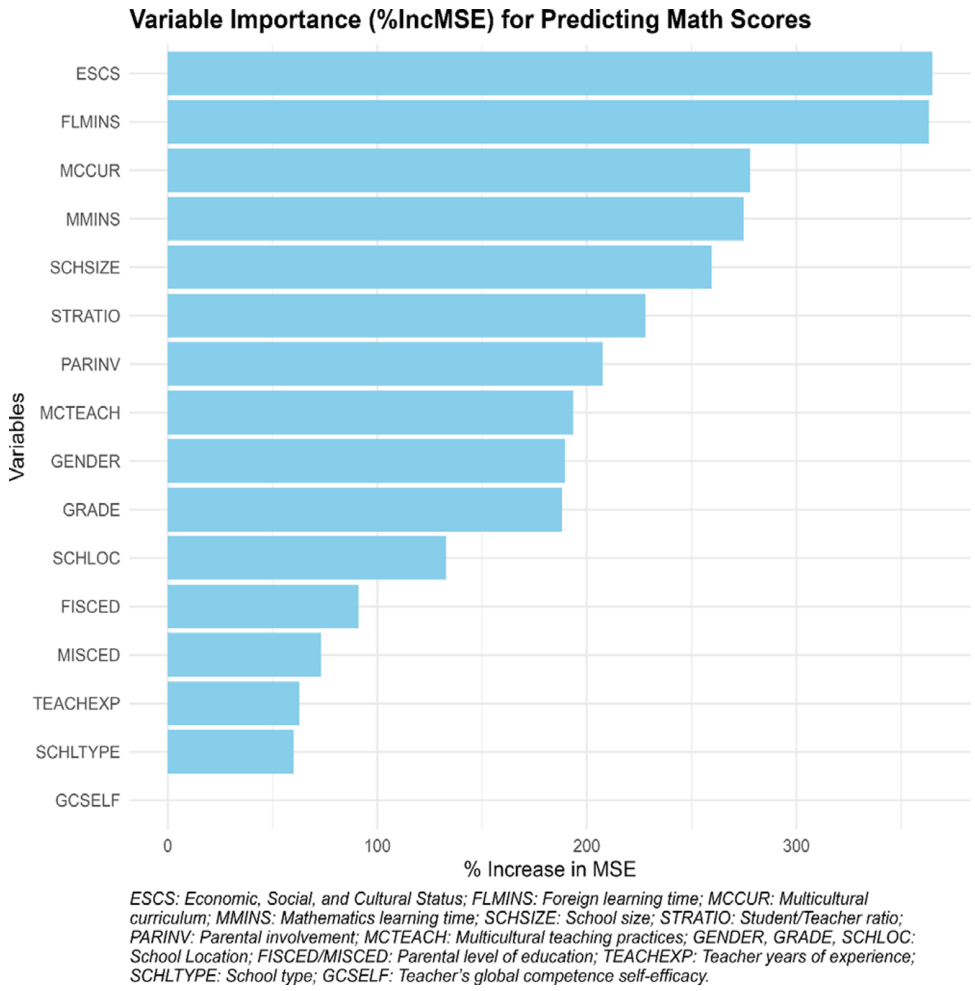


Figure 3. Random forest model results: Variable importance based on the percentage increase in mean-square error (%IncMSE – mean decrease in accuracy).

STRATIO (228.0%) demonstrated moderate importance. Notably, teachers' multicultural self-efficacy (*GCSELF*) had a % IncMSE of 0%, indicating that it had no predictive power for mathematics scores in this model. Figure 3 presents a ranked summary of predictor importance.

5.3. Robustness tests

5.3.1. Single PV value

A robustness check was conducted using a single PV (PV3). For sample sizes exceeding 6,500 cases, single-PV estimation yields stable results (OECD, 2009) and avoids the potential smoothing effect of averaging, ensuring the findings are not dependent on the aggregation procedure.

The PV3 model showed slightly lower performance ($R^2 = 0.962$, RMSE = 19.95, MAE = 10.83) than the averaged-PV model, but predictor importance patterns remained consistent, supporting the robustness of the findings.

5.3.2. Removal of PV-conditioning variables

PISA PVs are generated using latent proficiency models that incorporate some of the background variables included in our predictive models (e.g., socioeconomic status or *ESCS*, gender, grade) (Zieger et al., 2022). Consequently, it was deemed necessary to examine whether the predictive prominence of *FLMINS* was influenced by the inclusion of these variables. To do so, the RF model was re-estimated excluding *ESCS*, *GENDER*, and *GRADE*. As expected, model fit decreased to $R^2 = 0.83$, with RMSE = 36.92 and MAE = 23.73, corresponding to 7.6% and 4.9% of the mean mathematics score, respectively.

Despite this reduction in overall fit, in the absence of *ESCS*, *FLMINS* emerged as the most important predictor across both variable importance metrics (%IncMSE and IncNodePurity), surpassing all other student- and school-level variables. Other predictors retained their relative ranking, although their importance values changed modestly. These results demonstrate that the predictive contribution of *FLMINS* is not an artifact of PV conditioning and is robust across model specifications.

Full results, including detailed importance rankings and plots, are provided in the Supplementary Material (Appendix S2).

5.4. Study 2 discussion

Study 2 employed an RF model to examine associations between FL programme characteristics and mathematics achievement. The model explained a large share of variance ($R^2 = 0.97$, RMSE = 3.5%, MAE = 1.9% of mean score), with OOB performance closely matching test-set results, indicating strong internal consistency

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

and generalisability. However, the high level of model fit requires careful interpretation. PISA PVs are conditioned on background variables (such as SES, grade, and gender) that are also included as predictors – creating an outcome-predictor dependency known to inflate fit indices (Zieger et al., 2022). The large sample size and the presence of several highly predictive variables may have further contributed to this effect.

A robustness check using only PV3 produced a modestly lower R^2, consistent with the expected influence of pooling results across PVs, which smooths random error and can slightly increase fit. Crucially, when key PV-conditioning variables (*ESCS*, *GRADE*, and *GENDER*) were removed from the model, *FLMINS* emerged as the most important predictor across both importance metrics, and the ranking of the remaining variables remained highly stable. This pattern indicates that although absolute accuracy metrics are influenced by PV construction, the relative importance structure – particularly the prominence of *FLMINS* – reflects a genuine and robust signal rather than a modeling artifact.

Taken together, these findings suggest that the RF results should be interpreted primarily as exploratory, highlighting variable importance and nonlinear relationships rather than providing definitive estimates of predictive accuracy. Nonetheless, the stability of the importance rankings offers meaningful insights into the role of FL programme features within the broader ecosystem of predictors of mathematics achievement.

Consistent with Study 1, time dedicated to FL instruction (*FLMINS*) emerged as a highly influential predictor – second only to SES (*ESCS*) in the full model – a pattern that held across specifications. The RF model's capacity to capture nonlinearities and interactions further enriches this interpretation, offering a more nuanced understanding of how FL-related factors operate alongside other student- and school-level characteristics (Bates et al., 2015; Breiman, 2001).

Among the FL programme features, multicultural curriculum integration (*MCCUR*) demonstrated the strongest association with mathematics scores. This effect is likely mediated by the cognitive benefits of robust FL learning and cultural exposure, which is strongly linked to enhanced EFs – including working memory, response inhibition, and cognitive flexibility – all critical predictors of mathematical reasoning (Cragg et al., 2017; ten Braak et al., 2022). Additionally, multicultural education has been associated with enhanced critical thinking and problem-solving skills (Aslan & Aybek, 2020; Qondias et al., 2022), which also contribute to successful mathematical performance.

While this study offers credible evidence that features of FL programmes, particularly instructional time (*FLMINS*) and multicultural curriculum (*MCCUR*), are meaningfully associated with mathematical achievement, these findings do not establish causality. As in Study 1, additional research is required to disentangle the mechanisms underlying these associations.

6. General discussion

The present research examined the relationship between FL instruction time (*FLMINS*) and mathematics achievement among 15-year-old students using PISA 2018 data. Across all specifications, a consistent positive association was identified: Mathematics scores were on average 4.8 PISA points higher per additional hour of weekly FL instruction. Notably, this association is substantially stronger than that observed for mathematics instruction time (*MMINS*), which yielded a comparatively modest increment of 1.8 points per hour. This finding, which persisted after adjusting for powerful predictors like SES (*ESCS*) and grade level, underscores the unique and complementary contribution of FL learning to mathematical development.

6.1. Dual pathways and contextual moderation

The robustness of this positive association provides theoretical nuance, highlighting overlapping student-level and school-level components. The within–between decomposition showed a significant within-school effect: even when comparing peers in the same school, students with greater FL exposure tend to score higher in mathematics – consistent with our cognitive transfer theory. The stronger between-school effect suggests that unobserved institutional features may covary with FL time (e.g., curriculum, instructional quality, selection/admissions, timetabling).

These findings also engage with potential trade-offs in highly multilingual systems (Dentella et al., 2024). In a robustness check restricted to Luxembourg, Switzerland, Canada, Finland, Spain, and Singapore, the association persisted but was modest (3 points per additional FL hour/week). This pattern suggests that the link is sensitive to system-level features (curricular alignment, language-of-instruction policy), and learner-level factors (FL acquisition, usage), which may interact or influence cognitive development (Schiltz et al., 2024). Together, these considerations provide a framework through which our findings can coexist with prior reports of reduced maths achievement in contexts of high language burden, rather than contradict them.

6.2. Programme features and equity

Consistent with prior research, SES (*ESCS*) emerged as the strongest determinant of mathematics achievement (b = 29.33), highlighting persistent educational disparities linked to economic resources (Kim et al., 2019; OECD, 2023b; Rodríguez-Hernández et al., 2020). However, our findings suggest that FL study may serve as a potential avenue for promoting equity. Indeed, the magnitude of these associations is considerable: the predicted score difference associated with two hours of weekly FL instruction (9.6 points) is roughly equivalent to the observed gender gap, while a six-hour instructional volume (28.8 points) parallels the impact of a full SD increase in SES. Moreover, the positive association between *FLMINS* and maths achievement was observed across all economic status groupings, showing the strongest effect size in economies in transition. This suggests a compensatory role for FL programmes, where the strongest positive associations are observed in settings with fewer existing structural advantages.

The ML approach (Study 2) complemented this finding by confirming *FLMINS* as a strong predictive factor, comparable in importance to *ESCS* in predicting mathematical performance. Furthermore, multicultural curriculum integration (*MCCUR*) emerged as the most significant feature among the evaluated FL programme components, supporting the view that FL benefits may accrue not just from time on task, but from instruction delivered authentically, promoting higher-order thinking and culturally responsive learning environments.

6.3. Other determinants of achievement

Beyond the primary determinants, the models identified several conventional and anomalous predictors of mathematics achievement. Consistent with expectations, grade level exhibited a strong

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

effect (b = 28.08 per year of schooling), as students at higher grade levels have been exposed to more advanced topics and have had more opportunities to practice, apply concepts, and master fundamental skills.

Regarding institutional factors, school location (*SCHLOC*) and school type (*SCHLTYPE*) demonstrated moderate effects, indicating that students in urban settings and private schools achieve higher scores – a finding often linked to differential access to specialised resources and infrastructure. As expected, admission type (*ADMISSION*) also showed a significant positive effect, showing that selective or merit-based enrolment policies contribute to higher mean achievement. Institutional size factors, such as school size (*SCHSIZE*) and student-teacher ratio (*STRATIO*), were found to be of moderate predictive importance in Study 2, confirming that while these structural attributes may shape the learning environment, their impact on achievement is less pronounced than curricular content (Mohammadpour & Abdul Ghafar, 2014).

Our results also showed that boys outperformed girls by an average of 10.43 points, consistent with the documented gender gaps in mathematics performance (Anaya et al., 2022), which are likely influenced by contextual factors unrelated to biology, such as SES and cultural expectations (Breda et al., 2018; Johnson et al., 2022).

An interesting anomaly emerged regarding parental education (*MISCED*, *FISCED*). Contrary to typical findings (Davis-Kean et al., 2021), both maternal and paternal education showed a small negative association with scores. This finding is likely explained by the high correlation between parental education and the already-included *ESCS* variable ($r \approx 0.70$), suggesting that SES (*ESCS*) accounts for the primary variance. The residual, small negative effect might reflect subtle mechanisms, such as parental time constraints associated with highly demanding careers, or could simply be the result of multicollinearity within the structural model.

Finally, instructional time allocated to mathematics (*MMINS*) and teacher quality variables exhibited nuanced roles. The effect of increasing *MMINS* was nearly three times smaller than the association observed for *FLMINS* (1.8 versus 4.8 PISA points per hr/week), suggesting that gains from instructional time alone are incremental and likely secondary to other factors, such as pedagogical quality or student engagement. Correspondingly, teacher self-efficacy in multicultural environments (*GCSELF*), a predictor in Study 2, showed no predictive power for mathematics achievement. Contrary to previous studies (Abacioglu et al., 2020; Nuenay et al., 2024), this non-finding suggests a potential intention-action gap, where teacher confidence in managing diverse classrooms primarily impacts noncognitive outcomes (e.g., classroom climate) rather than translating directly into measurable gains in mathematics proficiency (Feng et al., 2024).

Collectively, these studies offer significant theoretical insights and suggest promising directions for policy. Our findings provide empirical support for the hypothesis that FL learning is a robust correlate of mathematical achievement. This relationship, established through a dual-pathway framework of cognitive transfer and institutional quality signaling, aligns with recent syntheses highlighting the potential of FL programmes to foster mathematical performance through relatively short instructional periods (Nucette et al., 2024).

Given the cross-sectional nature of the PISA data, these results are hypothesis generating for policy. They suggest that FL education should be viewed as a complementary component of a well-rounded curriculum, rather than a competitor for STEM hours. Notably, the 2.6-to-1 efficiency ratio – where FL instruction exhibits a significantly stronger association with maths scores than additional mathematics time – suggests that at the margin, prioritising cognitive breadth may be more effective for developing mathematical literacy than increasing subject-specific volume.

Furthermore, the consistent predictive role of *FLMINS* indicates it may serve as a lever for reducing academic disparities. Strategic FL investment offers a hypothesised "cognitive lift" capable of narrowing achievement gaps associated with gender and SES. Specifically, policy should prioritise integrating multicultural curricula to foster the abstract reasoning essential for numeracy and ensuring equitable access to high-quality FL programmes in under-resourced schools. However, it is crucial that any implementation of such initiatives consider the system-level constraints and potential trade-offs noted in multilingual contexts.

7. Limitations and future directions

While this research significantly expanded the scope and methodological rigour of the research on FL learning and mathematical achievement, several limitations warrant consideration. Mainly, the cross-sectional design prohibits any claims of causality, as unmeasured factors such as intrinsic motivation, teacher–student rapport, pedagogical approaches, or general cognitive ability may partly account for the observed association.

Additionally, the theoretical link to cognitive transfer mechanisms (e.g., EFs) remains indirect, as direct measures of these cognitive functions are unavailable in the PISA data.

Future research should adopt longitudinal or quasi-experimental designs to better establish causality, incorporate direct cognitive assessments to test mediation pathways, and conduct cross-cultural comparisons to clarify how system-level predictors (e.g., national policies, resource allocation, language regimes) and curricular contexts shape these relationships.

8. Conclusion

The results of these studies highlight the consistent and meaningful relationship between FL learning and mathematical achievement. By showing that the association between *FLMINS* and maths performance is robust across diverse contexts and driven by both individual and institutional factors, these findings reinforce the importance of providing access to high-quality FL education. Ultimately, the integration of FL instruction, particularly when accompanied by a multicultural curriculum, may present a promising strategy for supporting students' broader academic development and potentially mitigating contextual disparities in achievement.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1366728926101138.

Data availability statement. The data that support the findings of this study are openly available in the OECD's PISA 2018 Database (https://www.oecd.org/en/data/datasets/pisa-2018-database.html). The methodology details, along with the R code use for all models, are also available in the Supplementary Material (Appendices S3 and S4).

Acknowledgments. We would like to thank the two anonymous reviewers for their contributions to the previous version of the manuscript.

Competing interests. The authors declare none.

Disclosure of generative AI use. Generative artificial intelligence (AI) tools were used to assist with the troubleshooting of the R code for data analysis. Specifically, OpenAI's ChatGPT-4 model (April 2024 version) was utilised. No

AI tools were employed in the generation, interpretation, or critical analysis of the manuscript's scientific content. The authors accept full responsibility for the accuracy, integrity, and validity of all aspects of the work.

References

Abacioglu, C. S., **Volman, M.**, & **Fischer, A. H.** (2020). Teachers' multicultural attitudes and perspective taking abilities as factors in culturally responsive teaching. *British Journal of Educational Psychology*, **90**(3), 736–752. https://doi.org/10.1111/bjep.12328.

Abizada, A., & **Seyidova, S.** (2024). Effect of class size on student achievement at public secondary schools in Azerbaijan. *Cogent Education*, **11**(1), 2280306. https://doi.org/10.1080/2331186X.2023.2280306.

Al-khresheh, M., & **Karmi, S.** (2024). An exploration of cognitive benefits of EFL learning in a monolingual Jordanian context. *An-Najah University Journal for Research – B (Humanities)*, **38**(9), 1765–1794. https://doi.org/10.35552/0247.38.9.2250.

Anaya, L., **Stafford, F.**, & **Zamarro, G.** (2022). Gender gaps in math performance, perceived mathematical ability and college STEM education: The role of parental occupation. *Education Economics*, **30**(2), 113–128. https://doi.org/10.1080/09645292.2021.1974344.

Anderton, R., **Hine, G.**, & **Joyce, C.** (2017). Secondary school mathematics and science matters: Predicting academic success for secondary students transitioning into university allied health and science courses. *International Journal of Innovation in Science and Mathematics Education (IJISME)*, **25**(1). https://openjournals.library.sydney.edu.au/CAL/article/view/11317/11058

Arroyo Resino, D., **Constante-Amores, A.**, **Gil Madrona, P.**, & **Carrillo López, P. J.** (2024). Student well-being and mathematical literacy performance in PISA 2018: A machine-learning approach. *Educational Psychology*, **44**(3), 340–357. https://doi.org/10.1080/01443410.2024.2359104.

Aslan, S., & **Aybek, B.** (2020). Testing the effectiveness of interdisciplinary curriculum-based multicultural education on tolerance and critical thinking skill. *International Journal of Educational Methodology*, **6**(1), 43–55. https://doi.org/10.12973/ijem.6.1.43.

Bak, T. H., **Long, M. R.**, **Vega-Mendoza, M.**, & **Sorace, A.** (2016). Novelty, challenge, and practice: The impact of intensive language learning on attentional functions. *PLoS One*, **11**(4), e0153485. https://doi.org/10.1371/journal.pone.0153485.

Balabied, S. A. A., & **Eid, H. F.** (2023). Utilizing random forest algorithm for early detection of academic underperformance in open learning environments. *PeerJ Computer Science*, **9**, e1708. https://doi.org/10.7717/peerj-cs.1708.

Bates, D., **Mächler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bialystok, E., & **Barac, R.** (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, **122**(1), 67–73. https://doi.org/10.1016/j.cognition.2011.08.003.

Boonk, L., **Gijselaers, H. J. M.**, **Ritzen, H.**, & **Brand-Gruwel, S.** (2018). A review of the relationship between parental involvement indicators and academic achievement. *Educational Research Review*, **24**, 10–30. https://doi.org/10.1016/j.edurev.2018.02.001.

Breda, T., **Jouini, E.**, & **Napp, C.** (2018). Societal inequalities amplify gender gaps in math. *Science*, **359**(6381), 1219–1220. https://doi.org/10.1126/science.aar2307.

Breiman, L. (2001). Random forests. *Machine Learning*, **45**(1), 5–32. https://doi.org/10.1023/A:1010933404324.

Bull, R., & **Lee, K.** (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, **8**(1), 36–41. https://doi.org/10.1111/cdep.12059.

Byram, M. (2021). *Teaching and assessing intercultural communicative competence*. Multilingual Matters. https://doi.org/10.21832/9781800410251

Byram, M., **Holmes, P.**, & **Savvides, N.** (2013). Intercultural communicative competence in foreign language education: Questions of theory, practice and research. *The Language Learning Journal*, **41**(3), 251–253. https://doi.org/10.1080/09571736.2013.836343.

Chaika, O. (2023). Role of reflective practice in foreign language teaching and learning via multicultural education. *International Journal of Social Science and Human Research*, **6**(2), 1343–1350. https://doi.org/10.47191/ijsshr/v6-i2-74.

Chan, E. (2006). Teacher experiences of culture in the curriculum. *Journal of Curriculum Studies*, **38**(2), 161–176. https://doi.org/10.1080/00220270500391605.

Chiu, J., **Economos, J.**, **Markson, C.**, **Raicovi, V.**, **Howell, C.**, **Morote, E.**, & **Inserra, A.** (2016). Which matters Most? Perceptions of family income or parental education on academic achievement. *New York Journal of Student Affairs*, **16**(2), 3–16. https://touroscholar.touro.edu/gse_pubs/32

Choi, J. Y., **Jeon, S.**, & **Lippard, C.** (2018). Dual language learning, inhibitory control, and math achievement in head start and kindergarten. *Early Childhood Research Quarterly*, **42**, 66–78. https://doi.org/10.1016/j.ecresq.2017.09.001.

Cok, T. (2021). *Guidelines and recommendations for the development of cross-linguistic awareness for foreign language learning and teaching pathways to Plurilingual education* (pp. 111–129). University of Primorska Press. https://doi.org/10.26493/978-961-7055-36-8.111-129.

Colling, J., **Grund, A.**, **Keller, U.**, **Esch, P.**, **Fischbach, A.**, & **Ugen, S.** (2024). *Differences in Mathematics Achievement Between Students from European Public Schools and Students Following the Luxembourgish Curriculum: A cross-sectional analysis at primary and secondary school level (integral version).* Luxembourg Centre for Educational Testing (LUCET) & Service de la Recherche et de l'Innovation pédagogiques (SCRIPT). https://doi.org/10.48746/bb2024lu-en-33

Council of Europe. (2001). *Common European framework of reference for languages: learning, teaching, assessment (CEFR)* Cambridge University Press. https://www.coe.int/en/web/common-european-framework-reference-languages

Cragg, L., **Keeble, S.**, **Richardson, S.**, **Roome, H. E.**, & **Gilmore, C.** (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, **162**, 12–26. https://doi.org/10.1016/j.cognition.2017.01.014.

Crede, J., **Wirthwein, L.**, **McElvany, N.**, & **Steinmayr, R.** (2015). Adolescents' academic achievement and life satisfaction: The role of parents' education. *Frontiers in Psychology*, **6**. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2015.00052

Davis-Kean, P. E., **Tighe, L. A.**, & **Waters, N. E.** (2021). The role of parent educational attainment in parenting and children's development. *Current Directions in Psychological Science*, **30**(2), 186–192. https://doi.org/10.1177/0963721421993116.

Delaney, J. M., & **Devereux, P. J.** (2020). Math matters! The importance of mathematical and verbal skills for degree performance. *Economics Letters*, **186**, 108850. https://doi.org/10.1016/j.econlet.2019.108850.

Dentella, V., **Masullo, C.**, & **Leivada, E.** (2024). Bilingual disadvantages are systematically compensated by bilingual advantages across tasks and populations. *Scientific Reports*, **14**(1), 2107. https://doi.org/10.1038/s41598-024-52417-5.

Dixon, L. Q., **Zhao, J.**, **Shin, J.-Y.**, **Wu, S.**, **Su, J.-H.**, **Burgess-Brigham, R.**, **Gezer, M. U.**, & **Snow, C.** (2012). What we know about second language acquisition: A synthesis from four perspectives. *Review of Educational Research*, **82**(1), 5–60. https://doi.org/10.3102/0034654311433587.

Dlaska, A. (2000). Integrating culture and language learning in institution-wide language programmes. *Language, Culture and Curriculum*, **13**(3), 247–263. https://doi.org/10.1080/07908310008666602.

Eriksson, K., **Lindvall, J.**, **Helenius, O.**, & **Ryve, A.** (2021). Socioeconomic status as a multidimensional predictor of student achievement in 77 societies. *Frontiers in Education*, **6**. https://doi.org/10.3389/feduc.2021.731634.

Esposito, A. G. (2020). Executive functions in two-way dual-language education: A mechanism for academic performance. *Bilingual Research Journal*, **43**(4), 417–432. https://doi.org/10.1080/15235882.2021.1874570.

European Commission. (2011). *Mathematics education in Europe – Common challenges and national policies*. Publications Office. https://doi.org/10.2797/72660.

Feng, X., **Zhang, N.**, **Yang, D.**, **Lin, W.**, & **Maulana, R.** (2024). From awareness to action: Multicultural attitudes and differentiated instruction of teachers in Chinese teacher education programmes. *Learning Environments Research*. https://doi.org/10.1007/s10984-024-09518-9.

Fiskerstrand, A. (2022). Literature review – Parent involvement and mathematic outcome. *Educational Research Review*, **37**, 100458. https://doi.org/10.1016/j.edurev.2022.100458.

Forgasz, H. J., & **Hill, J. C.** (2013). Factors implicated in high mathematics achievement. *International Journal of Science and Mathematics Education*, **11**(2), 481–499. https://doi.org/10.1007/s10763-012-9348-x.

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

Freeman, B., **Marginson, S.**, & **Tytler, R.** (2019). An international view of STEM education. In *STEM education 2.0* (pp. 350–363). Brill. https://doi.org/10.1163/9789004405400_019.

Fuhs, M. W., **Nesbitt, K. T.**, **Farran, D. C.**, & **Dong, N.** (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology*, **50**(6), 1698–1709. https://doi.org/10.1037/a0036633.

Gelman, A., & **Hill, J.** (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790942.

Gil-Madrona, P., **Guerrero-Muguerza, A. M.**, **Infantes-Paniagua, Á.**, & **Martínez-López, M.** (2025). What are the best predictors of STEM competences in PISA 2018? An analysis of the Spanish context using data mining. *School Science and Mathematics*. https://doi.org/10.1111/ssm.18363.

González-Martín, A. M., **Berd-Gómez, R.**, **Saura-Montesinos, V.**, **Biel-Maeso, M.**, & **Abrahamse, E.** (2024). On the relationship between bilingualism and mathematical performance: A systematic review. *Education Sciences*, **14**(11). https://doi.org/10.3390/educsci14111172.

Greisen, M., **Georges, C.**, **Hornung, C.**, **Sonnleitner, P.**, & **Schiltz, C.** (2021). Learning mathematics with shackles: How lower reading comprehension in the language of mathematics instruction accounts for lower mathematics achievement in speakers of different home languages. *Acta Psychologica*, **221**, 103456. https://doi.org/10.1016/j.actpsy.2021.103456.

Hamamura, T. (2011). Power distance predicts gender differences in math performance across societies. *Social Psychological and Personality Science*, **3**(5), 545–548. https://doi.org/10.1177/1948550611429191.

Hanushek, E., & **Woessmann, L.** (2017). School resources and student achievement: A review of cross-country economic research. In M. Rosén, K. Y. Hansen, & U. Wolff (Eds.), *Cognitive abilities and educational outcomes* (1st ed., pp. 149–171). Springer Cham. https://doi.org/10.1007/978-3-319-43473-5_8.

Hanushek, E. A., & **Woessmann, L.** (2021). *Education and economic growth Oxford research Encyclopedia of economics and finance*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190625979.013.651.

Hastie, T., **Tibshirani, R.**, & **Friedman, J.** (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Hong, S., **Yoo, S.-K.**, **You, S.**, & **Wu, C.-C.** (2010). The reciprocal relationship between parental involvement and mathematics achievement: Autoregressive cross-lagged Modeling. *The Journal of Experimental Education*, **78**(4), 419–439. https://doi.org/10.1080/00220970903292926.

Huang, T., **Loerts, H.**, & **Steinkrauss, R.** (2022). The impact of second- and third-language learning on language aptitude and working memory. *International Journal of Bilingual Education and Bilingualism*, **25**(2), 522–538. https://doi.org/10.1080/13670050.2019.1703894.

Huisman, J., **Smolentseva, A.**, & **Froumin, I.** (Eds.). (2018). *25 years of transformations of higher education Systems in Post-Soviet Countries*. Springer Nature. https://doi.org/10.1007/978-3-319-52980-6.

Hyde, J. S., & **Mertz, J. E.** (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences*, **106**(22), 8801–8807. https://doi.org/10.1073/pnas.0901265106.

Iglesias-Sarmiento, V., **Carriedo, N.**, **Rodríguez-Villagra, O. A.**, & **Pérez, L.** (2023). Executive functioning skills and (low) math achievement in primary and secondary school. *Journal of Experimental Child Psychology*, **235**, 105715. https://doi.org/10.1016/j.jecp.2023.105715.

James, G., **Witten, D.**, **Hastie, T.**, & **Tibshirani, R.** (2023). *An introduction to statistical learning with applications in R second edition*. Springer https://hastie.su.domains/ISLR2/ISLRv2_corrected_June_2023.pdf.download.html

Jewsbury, P. A., **Jia, Y.**, & **Gonzalez, E. J.** (2024). Considerations for the use of plausible values in large-scale assessments. *Large-scale Assessments in Education*, **12**(1), 24. https://doi.org/10.1186/s40536-024-00213-y.

Joensen, J. S., & **Nielsen, H. S.** (2009). Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources*, **44**(1), 171–198. https://doi.org/10.3368/jhr.44.1.171.

Johnson, T., **Burgoyne, A. P.**, **Mix, K. S.**, **Young, C. J.**, & **Levine, S. C.** (2022). Spatial and mathematics skills: Similarities and differences related to age, SES, and gender. *Cognition*, **218**, 104918. https://doi.org/10.1016/j.cognition.2021.104918.

Kim, S., **Cho, H.**, & **Kim, L. Y.** (2019). Socioeconomic status and academic outcomes in developing countries: A meta-analysis. *Review of Educational Research*, **89**(6), 875–916. https://doi.org/10.3102/0034654319877155.

Lavrenteva, E., & **Orland-Barak, L.** (2015). The treatment of culture in the foreign language curriculum: An analysis of national curriculum documents. *Journal of Curriculum Studies*, **47**(5), 653–684. https://doi.org/10.1080/00220272.2015.1056233.

Lee, H. (2010). *The relationship between Korean EFL high school students' English achievement and other subject achievement*. [Doctor of Education, Alliant International University]. ProQuest Dissertations & Theses. https://www.proquest.com/docview/863582967

Lee Swanson, H., **Arizmendi, G. D.**, & **Li, J.-T.** (2021). Working memory growth predicts mathematical problem-solving growth among emergent bilingual children. *Journal of Experimental Child Psychology*, **201**, 104988. https://doi.org/10.1016/j.jecp.2020.104988.

Liaw, A., & **Wiener, M.** (2002). Classification and regression by randomForest. *R News*, **2**, 18–22. https://CRAN.R-project.org/doc/Rnews/

Liew, K., **Hamamura, T.**, & **Uchida, Y.** (2025). Machine learning culture: Cultural membership classification as an exploratory approach to cross-cultural psychology. *Personality and Social Psychology Bulletin*, 01461672251339313. https://doi.org/10.1177/01461672251339313.

Lin, M., **Lucas, H. C.**, & **Shmueli, G.** (2013). Research commentary: Too big to fail: Large samples and the p-value problem. *Information Systems Research*, **24**(4), 906–917. http://www.jstor.org/stable/24700283

Lüdecke, D., **Ben-Shachar, M. S.**, **Patil, I.**, **Waggoner, P.**, & **Makowski, D.** (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, **6**(60), 3139. https://doi.org/10.21105/joss.03139.

Lusin, N., **Peterson, T.**, **Sulewski, C.**, & **Zafer, R.** (2023). *Enrollments in languages other than English in US Institutions of Higher Education, Fall 2021*. https://www.mla.org/content/download/191324/file/Enrollments-in-Languages-Other-Than-English-in-US-Institutions-of-Higher-Education-Fall-2021.pdf

Mohammadpour, E., & **Abdul Ghafar, M. N.** (2014). Mathematics achievement as a function of within- and between-school differences. *Scandinavian Journal of Educational Research*, **58**(2), 189–221. https://doi.org/10.1080/00313831.2012.725097.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, **46**(1), 69–85. https://doi.org/10.2307/1913646.

Nachouki, M., **Mohamed, E. A.**, **Mehdi, R.**, & **Abou Naaj, M.** (2023). Student course grade prediction using the random forest algorithm: Analysis of predictors' importance. *Trends in Neuroscience and Education*, **33**, 100214. https://doi.org/10.1016/j.tine.2023.100214.

National Standards. (2015). *World-readiness standards for learning languages*. National Standards in Foreign Language Education Project. https://www.actfl.org/uploads/files/general/World-ReadinessStandardsforLearningLanguages.pdf

Nollenberger, N., **Rodríguez-Planas, N.**, & **Sevilla, A.** (2016). The math gender gap: The role of culture. *American Economic Review*, **106**(5), 257–261. https://doi.org/10.1257/aer.p20161121.

Nucette, A., **Hamamura, T.**, **Leitao, S.**, & **Biedermann, B.** (2024). Can learning a new language make you better at maths? A meta-analysis of foreign language learning and numeracy skills during early adolescence. *Bilingualism: Language and Cognition*, 1–31. https://doi.org/10.1017/S1366728924000701.

Nuenay, E., **Cariga, T.**, **Bualan, G.**, & **Banes, C.** (2024). Teacher's attitudes and efficacy in teaching multicultural classrooms. *International Journal of Research and Innovation in Social Science*, **VIII**, 4783–4795. https://doi.org/10.47772/IJRISS.2024.8080364.

OECD. (2009). *PISA data analysis manual: SPSS* (2nd ed.). https://doi.org/10.1787/9789264056275-en.

OECD. (2019a). *PISA 2018 Assessment and Analytical Framework*. https://doi.org/10.1787/b25efab8-en

OECD. (2019b). *PISA 2018 Results (Volume I): What Students Know and Can Do*. https://doi.org/10.1787/5f07c754-en

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press

OECD. (2020). Annex a: PISA 2018 technical background. In *PISA 2018 results (volume VI): Are students ready to thrive in an interconnected world?* OECD Publishing. https://doi.org/10.1787/d5f68679-en.

OECD. (2023a). *Education at a Glance 2023: OECD Indicators.* https://doi.org/10.1787/e13bef63-en

OECD. (2023b). *PISA 2022 results (volume I): The state of learning and equity in education*. https://doi.org/10.1787/53f23881-en.

OECD. (2024). *Programme for International Student Assessment (PISA).* https://www.oecd.org/pisa/

Olivoto, T., & **Lúcio, A. D.** (2020). Metan: An R package for multi-environment trial analysis. *Methods in Ecology and Evolution*, **11**(6), 783–789. https://doi.org/10.1111/2041-210X.13384.

Olson, J. C., **Cooper, S.**, & **Lougheed, T.** (2011). Influences of teaching approaches and class size on undergraduate mathematical learning. *Primus*, **21**(8), 732–751. https://doi.org/10.1080/10511971003699694.

Padilla, A. M., **Fan, L.**, **Xu, X.**, & **Silva, D.** (2013). A mandarin/English two-way immersion program: Language proficiency and academic achievement. *Foreign Language Annals*, **46**(4), 661–679. https://doi.org/10.1111/flan.12060.

Park, S., **Dotan, P. L.**, & **Esposito, A. G.** (2023). Do executive functions gained through two-way dual-language education translate into math achievement? *International Journal of Bilingual Education and Bilingualism*, **26**(4), 457–471. https://doi.org/10.1080/13670050.2022.2116973.

Pishghadam, R., & **Zabihi, R.** (2011). Parental education and social and cultural capital in academic achievement. *International Journal of English Linguistics*, **1**(2), 50. https://doi.org/10.5539/ijel.v1n2p50.

Qondias, D., **Lasmawan, W.**, **Dantes, N.**, & **Arnyana, I. B. P.** (2022). Effectiveness of multicultural problem-based learning models in improving social attitudes and critical thinking skills of elementary school students in thematic instruction. *Journal of Education and e-Learning Research*, **9**(2), 62–70. https://doi.org/10.20448/jeelr.v9i2.3812.

Ramos, C., **Jadán-Guerrero, J.**, & **Gómez-García, A.** (2018). Relationship between academic performance and the self-report of the executive performance of Ecuadorian teenagers. *Avances en Psicología Latinoamericana*, **36**(2), 405–417. https://doi.org/10.12804/revistas.urosario.edu.co/apl/a.5481.

Rodríguez, S., **Piñeiro, I.**, **Gómez-Taibo, M.**, **Regueiro, B.**, **Estévez, I.**, & **Valle, A.** (2017). An explanatory model of maths achievement: Perceived parental involvement and academic motivation. *Psicothema*, **29**(2), 184–190. https://doi.org/10.7334/psicothema2017.32.

Rodríguez-Hernández, C. F., **Cascallar, E.**, & **Kyndt, E.** (2020). Socio-economic status and academic performance in higher education: A systematic review. *Educational Research Review*, **29**, 100305. https://doi.org/10.1016/j.edurev.2019.100305.

Ryshina-Pankova, M. (2015). Foreign language curriculum as a means of achieving humanities learning goals: Assessment of materials, pedagogy, and learner texts. In J. M. Norris & J. M. Davis (Eds.), *Student learning outcomes assessment in college foreign language programs* (pp. 221–246). National Foreign Language Resource Center.

Sabine, P. (2020). *The development of executive functions in childhood and adolescence and their relation to school performance* (1st ed.). Routledge. https://doi.org/10.4324/9781003016830-13.

Savignon, S. J., & **Sysoyev, P. V.** (2005). Cultures and comparisons: Strategies for learners. *Foreign Language Annals*, **38**(3), 357–365. https://doi.org/10.1111/j.1944-9720.2005.tb02222.x.

Schiltz, C., **Lachelin, R.**, **Hilger, V.**, & **Marinova, M.** (2024). Thinking about numbers in different tongues: An overview of the influences of multilingualism on numerical and mathematical competencies. *Psychological Research*, **88**(8), 2416–2431. https://doi.org/10.1007/s00426-024-01997-y.

Shoghi, S., & **Ghonsooly, B.** (2018). Learning a foreign language: A new path to enhancement of cognitive functions. *Journal of Psycholinguistic Research*, **47**(1), 125–138. https://doi.org/10.1007/s10936-017-9518-7.

Silinskas, G., & **Kikas, E.** (2019). Parental involvement in math homework: Links to children's performance and motivation. *Scandinavian Journal of Educational Research*, **63**(1), 17–37. https://doi.org/10.1080/00313831.2017.1324901.

Song, Y., & **Cutumisu, M.** (2025). Using machine learning to predict student science achievement based on science curriculum type in TIMSS 2019. *International Journal of Science Education*, **47**(9), 1105–1149. https://doi.org/10.1080/09500693.2024.2359099.

Stoet, G., & **Geary, D. C.** (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, **29**(4), 581–593. https://doi.org/10.1177/0956797617741719.

Sullivan, M. D., **Janus, M.**, **Moreno, S.**, **Astheimer, L.**, & **Bialystok, E.** (2014). Early stage second-language learning improves executive control: Evidence from ERP. *Brain and Language*, **139**, 84–98. https://doi.org/10.1016/j.bandl.2014.10.004.

Tabachnick, B. G., & **Fidell, L. S.** (2018). *Using multivariate statistics* (7th ed.). Pearson.

ten Braak, D., **Lenes, R.**, **Purpura, D. J.**, **Schmitt, S. A.**, & **Størksen, I.** (2022). Why do early mathematics skills predict later mathematics and reading achievement? The role of executive function. *Journal of Experimental Child Psychology*, **214**, 105306. https://doi.org/10.1016/j.jecp.2021.105306.

Thiese, M. S., **Ronna, B.**, & **Ott, U.** (2016). P value interpretations and considerations. *Journal of Thoracic Disease*, **8**(9), E928–e931. https://doi.org/10.21037/jtd.2016.08.16.

UNESCO. (2021). *Using ISCED diagrams to compare education systems.* UNESCO Institute for Statistics. https://isced.uis.unesco.org/wp-content/uploads/sites/15/2021/07/UIS-ISCED-DiagramsCompare-OECDAnnex-final.pdf

United Nations. (2025). *World Economic Situation and Prospects 2025.* https://www.developmentaid.org/api/frontend/cms/file/2025/01/WESP-2025_Official_WEB.pdf

Valero, A. (2021). *Education and economic growth*. London School of Economics and Political Science. https://cep.lse.ac.uk/pubs/download/dp1764.pdf

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, **28**(2), 3–28. https://doi.org/10.1257/jep.28.2.3.

Vega-Mendoza, M., **West, H.**, **Sorace, A.**, & **Bak, T. H.** (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition*, **137**, 40–46. https://doi.org/10.1016/j.cognition.2014.12.008.

Watzinger-Tharp, J., **Swenson, K.**, & **Mayne, Z.** (2016). Academic achievement of students in dual language immersion. *International Journal of Bilingual Education and Bilingualism*, **21**, 1–16. https://doi.org/10.1080/13670050.2016.1214675.

White, L. J., & **Greenfield, D. B.** (2017). Executive functioning in Spanish- and English-speaking head start preschoolers. *Developmental Science*, **20**(1). https://doi.org/10.1111/desc.12502.

Wilcox, R. (2017). *Introduction to robust estimation and hypothesis testing* (4th ed.). Academic Press. https://doi.org/10.1016/B978-0-12-804733-0.00018-4.

Woessmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *The Journal of Economic Perspectives*, **30**(3), 3–31. http://www.jstor.org/stable/43855699

Woll, B., & **Wei, L.** (2019). *Cognitive Benefits of Language Learning: Broadening our perspectives. Final Report to the British Academy.* https://www.thebritishacademy.ac.uk/documents/287/Cognitive-Benefits-Language-Learning-Final-Report.pdf

Wong, P. C. M., **Ou, J.**, **Pang, C. W. Y.**, **Zhang, L.**, **Chi, S. T.**, **Lam, L. C. W.**, & **Antoniou, M.** (2019). Language training leads to global cognitive improvement in older adults: A preliminary study. *Journal of Speech, Language, and Hearing Research*, **62**(7), 2411–2424. https://doi.org/10.1044/2019_JSLHR-L-18-0321.

Wood, S. N. (2017). *Generalized additive models: An introduction with R, second edition*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279.

Xu, C., **Di Lonardo Burr, S.**, **Skwarchuk, S.-L.**, **Douglas, H.**, **Lafay, A.**, **Osana, H. P.**, **Simms, V.**, **Wylie, J.**, **Maloney, E. A.**, & **LeFevre, J.-A.** (2022). Pathways to learning mathematics for students in French-immersion and English-instruction programs. *Journal of Educational Psychology*, **114**(6), 1321–1342. https://doi.org/10.1037/edu0000722.

Yarkoni, T., & **Westfall, J.** (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, **12**(6), 1100–1122. https://doi.org/10.1177/1745691617693393.

Zieger, L. R., **Jerrim, J.**, **Anders, J.**, & **Shure, N.** (2022). Conditioning: How background variables can influence PISA scores. *Assessment in Education: Principles, Policy & Practice*, **29**(6), 632–652. https://doi.org/10.1080/0969594X.2022.2118665.

https://doi.org/10.1017/S1366728926101138 Published online by Cambridge University Press