Studies in Educational Evaluation 90 (2026) 101620

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

Understanding disparities: Student and school factors associated with U.S. students' achievement in reading, mathematics, and science

Nirmal Ghimire [*],[1]

Watson College of Education, University of North Carolina Wilmington, United States

ARTICLE INFO

Keywords:
Multilevel modeling
Educational disparities
Ecological system theory
Post-pandemic education
Cross-subject comparison
Compositional effects

ABSTRACT

This study examines multilevel associations between student characteristics, school contexts, and academic achievement across reading, mathematics, and science using PISA 2022 U.S. data. Hierarchical linear models with 4552 students nested within 154 schools reveal substantial between-school variance, highest for mathematics (24.5%), followed by science (22.4%) and reading (18.9%). Each of the 10 plausible values per domain was treated as an imputation of the latent achievement construct, fit on every PV x covariate-imputation combination (100 fits per model), and pooled via Rubin's rules. Gender demonstrates domain-specific patterns: males score 22 points lower in reading but 13 points higher in mathematics, with marginal advantage in science. Home language and parental education show consistent positive associations across domains. School-mean parental education and home ICT resources are associated with roughly 3.7 and 11 times the corresponding within-school effects. Cross-level interactions are significant though random-slopes reveal substantial school-level heterogeneity in demographic associations.

1. Introduction

The most recent Programme for International Student Assessment (PISA) results reveal concerning patterns in U.S. student performance, with mathematics scores showing a difference of 13 points compared to pre-pandemic levels, while reading and science scores remained relatively stable despite global disruptions (OECD, 2023; Schleicher, 2023). These varying patterns across subject domains raise important questions about how different academic subjects may be differentially affected by educational disruptions and whether these patterns hold true across different student populations and school contexts. This study aims to investigate whether these broad national trends remain consistent when accounting for the multilevel nature of educational systems and the diverse characteristics of students and schools.

PISA 2022 is among the first international assessments administered after the COVID−19 disruptions, during which school closures affected approximately 1.6 billion learners globally and learning losses were most pronounced in mathematics and among socioeconomically disadvantaged students (Donnelly & Patrinos, 2021; Maldonado & De Witte, 2022; Meinck et al., 2022). In the United States, mathematics scores declined while reading and science remained relatively stable (Gajderowicz et al., 2025; Sparks, 2023), motivating questions about whether and how multilevel patterns of achievement vary across subject domains in the post-pandemic context.

1.1. Research gap and original contribution

Despite extensive research on educational achievement using PISA data, three Key limitations characterize much of the existing literature. First, many analyses fail to account for the hierarchical nature of the educational data, where students are nested within schools. Traditional single-level analyses can produce biased standard errors and incorrect statistical inferences when applied to clustered data, particularly when intraclass correlations are substantial (Raudenbush & Bryk, 2002). Second, most studies often examine individual subjects in isolation rather than systematically comparing patterns across all three core PISA domains within a single analytical framework (Kuhfeld et al., 2022). This domain-by-domain approach obscures potential insights about whether background factors show consistent effects across subjects or demonstrate domain-specific patterns. Third, existing research has not

---

* Correspondence to: TESL Program, Education Building (EB345), Watson College of Education, University of North Carolina Wilmington, 601 South College Road, Wilmington, NC 28403, United States.

E-mail address: ghimiren@uncw.edu.

[1] ORCID: http://orcid.org/0000-0002-2032-1624

https://doi.org/10.1016/j.stueduc.2026.101620
Received 14 April 2025; Received in revised form 26 April 2026; Accepted 25 May 2026
Available online 3 June 2026

yet comprehensively examined the post-pandemic context captured in PISA 2022, particularly regarding how multilevel factors may relate differently to achievement across subject domains following significant educational disruptions.

This study addresses these gaps through four distinct contributions that extend beyond simply analyzing recent data. First, it employs hierarchical linear modeling approach that simultaneously examines variance decomposition, contextual effects, compositional effects, and random slopes across all three subject domains. This integrated analytical strategy reveals patterns that single-model approach cannot detect, particularly regarding how the strength of individual-level relationships vary systematically across school contexts. The random slopes analysis provides an important methodological extension by quantifying institutional heterogeneity in demographic achievement relationships, demonstrating that disparities are not uniform but vary substantially depending on school contexts. This finding has important implications for educational policy, suggesting that some schools successfully mitigate demographic achievement gaps while others amplify them.

Second, the study provides systematic cross domain comparison that identifies both universal and subject-specific patterns in how hierarchical factors relate to achievement. While some research examines multiple subjects, few studies conduct parallel analyses with identical model specifications across reading, mathematics, and science within the same sample.

Third, the analysis explicitly examines compositional effects, distinguishing between individual-level and school aggregate effects of parental education and technological resources. This methodological approach addresses a longstanding debate in stratification research about whether school composition effects represent true contextual influences or merely artifacts of aggregated individual characteristics (Perry & McConney, 2010; Willms, 2010). The findings reveal compositional relationships of measured variables, and the patterns illuminate mechanisms through which educational inequalities may be amplified through institutional sorting and peer effects.

Fourth, by focusing specifically on the United States, the study provides detailed insights into achievement patterns within a particular educational system characterized by substantial school autonomy, local funding mechanisms, and high levels of between-school segregation by socioeconomic status. While single-country analyses cannot test cross-national generalizability, they offer important advantages for understanding context-specific mechanisms and informing domestic policy discussions. The U.S. context is particularly relevant given ongoing debates about educational equity, school funding reform, and strategies for addressing pandemic-related learning disruptions. Moreover, the methodological approaches demonstrated here can be replicated in other national contexts, enabling future comparative research to examine whether the observed patterns reflect universal processes or country-specific dynamics.

Specifically, this study employs hierarchical linear modeling to examine how student and school characteristics relate to achievement in reading, mathematics, and science while accounting for the nested structures of educational environments. The following research questions guide the analysis:

1. What proportion of variance in reading, mathematics, and science achievement among U.S. 15-year-olds in PISA 2022 is attributable to student-level versus school-level factors, and does this variance decomposition differ across subject domains?
2. How do individual background factors (gender, immigration status, parental education, home language, and technological resources) differently relate to performance across reading, mathematics, and science domains in the post-pandemic context?
3. To what extent do school socioeconomic composition and community type moderate the relationships between student demographic characteristics and achievement outcomes, and do they vary across reading, mathematics, and science domains?

2. Literature review

2.1. Theoretical framework

Educational research increasingly recognizes that student achievement develops within interrelated environmental systems rather than resulting from isolated factors (Lee & Burkam, 2002). Bronfenbrenner's ecological system theory provides a useful framework for understanding these complex influences, particularly when examining academic outcomes across multiple subject domains (Bronfenbrenner & Morris, 2007). This theoretical perspective posits that development occurs through progressive interactions between individuals and their immediate environments (microsystems), connections between settings (mesosystems), external influences (exosystems), broader social contexts (macrosystems), and changes over time (chronosystems).

In the present study, the framework translates as follows (Wentzel & Ramani, 2016): individual factors (gender, immigration status, parental education, home language, home ICT resources) operate at the microsystem level; school-aggregate parental education and ICT resources reflect mesosystem connections between home and school; school community type, socioeconomic disadvantage concentration, and institutional ICT resources represent the exosystem; and the COVID−19 pandemic constitutes a chronosystem disruption that potentially altered relationships across the other levels. The inclusion of ICT at both levels aligns with Digital Divide scholarship on how technology access intersects with traditional socioeconomic factors (Chiu, 2020; van Dijk, 2005), with home ICT representing household capacity for technology-mediated learning and school ICT representing institutional capacity for digitally-enhanced instruction.

This ecological framework aligns with contemporary educational effectiveness research, including the Dynamic Model of Educational Effectiveness (Kyriakides et al., 2020). Multilevel modeling operationalizes this ecological perspective by accounting for the nested nature of educational data (Raudenbush & Bryk, 2002).

PISA 2022 provides a unique opportunity to examine these relationships during a period of significant educational upheaval (Schleicher, 2023), during which the pandemic magnified existing inequalities, students with fewer home technology resources or less-educated parents faced greater remote-learning challenges, while school-level digital infrastructure may have buffered or exacerbated these vulnerabilities (Donnelly & Patrinos, 2021).

This study uses Bronfenbrenner's framework to examine how individual and school factors relate to reading, mathematics, and science achievement in the post-pandemic context. By employing hierarchical modeling with PISA 2022 data, the analysis can identify whether ecological factors show consistent patterns across domains or demonstrate subject-specific relationships, providing insights into how educational systems might address the differential impacts of educational disruptions.

2.2. Individual background factors and achievement

Recent research reveals complex patterns in how demographic, socioeconomic, and technological resources at the individual level correlate with performance across reading, mathematics, and science domains. Within Bronfenbrenner's ecological framework, these individual characteristics represent microsystem influences, factors operating within students' immediate environments that directly shape their developmental experiences and learning opportunities.

2.2.1. Gender differences across subject domains

Gender represents a fundamental individual characteristic that consistently demonstrates domain-specific associations with academic achievement. Analysis of PISA results consistently shows that boys tend to perform better in mathematics while girls demonstrate stronger reading performance (Bay & Monseur, 2016). The PISA 2022 results confirm these patterns, with boys outperforming girls in mathematics by nine score points on average across OECD countries, while girls outperformed boys in reading by 24 score points (OECD, 2023).

The gender-based achievement patterns show remarkable consistency across countries and over time. Research examining PISA data across multiple cycles found that these gender differences in mathematics and reading are inversely related, i.e., countries with smaller gender gaps in mathematics tend to have larger gaps in reading (Stoet & Geary, 2013). This suggests that subject-specific gender differences may reflect distinct cognitive or educational processes rather than overall academic capability.

2.2.2. Immigration background and home language

Immigration status and language background represent critical individual factors affecting student achievement, operating at the microsystem level through direct influences on students' linguistic development and cultural integration experiences. Research using PISA data has consistently shown that immigrant students, particularly first-generation immigrants, tend to perform lower than their native-born peers across subject domains (OECD, 2018). This pattern is especially pronounced in reading, where language proficiency plays a critical role.

The relationship between immigrant status and achievement is moderated by several factors. Language spoken at home emerges as particularly significant, with students who speak the test language at home performing substantially better than those who don't. In many countries, "students who speak the language of instruction at home scored higher in reading than those who do not" with differences exceeding 50 score points (OECD, 2018).

Recent research continues to underscore the complex relationship between immigration status and academic achievement. Camacho and Fuligni (2015) found that first-generation immigrant students exhibit particularly pronounced deficits in reading compared to mathematics and science, highlighting the linguistic barriers inherent in the immigration-achievement nexus. Expanding on this, Ghimire (2024) demonstrated that both generational status and the language spoken at home significantly influence student outcomes across all core subjects. Native students significantly outperformed first-generation immigrants in reading, math, and science, while no statistically significant differences emerged between native and second-generation students. Similarly, students who speak the test language (English) at home demonstrated markedly higher achievement than those who speak Spanish or other languages, with gaps particularly pronounced in reading.

2.2.3. Socioeconomic status and parental education

Parental education serves as a key microsystem-level indicator of socioeconomic status, reflecting the educational resources, academic expectations, and learning support available within students' immediate home environments. Extensive research demonstrates robust positive associations between parental education and student achievement across all subject domains. Studies using PISA data consistently find that students whose parents have higher educational attainment show better academic outcomes, with effects evident across reading, mathematics, and science (Eriksson et al., 2021; Ghimire & Mokhtari, 2025).

The mechanisms through which parental education influences achievement are multifaceted. Higher parental education typically correlates with greater access to educational resources, more frequent engagement in academic activities, higher educational expectations, and enhanced capacity to support children's learning (Xie & Ma, 2019). These advantages accumulate over time, contributing to higher achievement among children from parents with high educational attainment.

Research examining multiple dimensions of socioeconomic status suggests that parental education may be particularly influential in shaping educational outcomes. Avvisati (2020) reviewed the measurement of socioeconomic status in PISA, noting the central role of parental education and home possessions in the ESCS index, while empirical analyses demonstrate that these components show strong the relatively consistent associations with achievement across countries, though magnitudes very depending on educational system and social policies (Eriksson et al., 2021).

The COVID−19 pandemic potentially amplified the importance of parental education, as remote learning placed greater demands on families to support their children's education. Parents with higher education levels may have been better positioned to provide academic assistance, navigate online learning platforms, and maintaining educational routines during school closures.

2.2.4. Information and communication technology (ICT Resources)

Access to information and communication technology (ICT) resources at home represents an increasingly critical microsystem-level factor, particularly following the shift to remote learning during the COVID−19 pandemic.

Analysis of PISA data indicates that basic access to computers and internet connectivity relates positively to achievement, but the relationship is not simply linear. Chiu (2020) found that moderate levels of ICT use were associated with higher reading performance, while excessive use showed diminishing returns or even negative associations. Ghimire and Mokhtari (2025) extended this work using structural equation modeling: basic digital access (internet connectivity, device ownership), supported learning outcomes, but an overabundance of devices showed counterproductive returns, a pattern consistent with technology saturation effects reported elsewhere (Borgonovi & Pokropek, 2021; Gubbels et al., 2020). These curvilinear patterns coexist with persistent equity concerns, including the "second digital divide", i.e., qualitative differences in how technology is used for educational purposes across socioeconomic groups (van Dijk, 2005).

2.3. School contextual factors and achievement

Research consistently demonstrates that school environments significantly contribute to academic outcomes beyond individual student factors (Wang et al., 2023). Within Bronfenbrenner's ecological framework, school-level characteristics operate at the mesosystem and exosystem levels, representing both the connections between students' immediate environments (home and school) and the broader institutional contexts that indirectly shape their developmental experiences. School socioeconomic composition reflects mesosystem dynamics, capturing how the collective characteristics of students and families create distinct peer and community contexts within schools. School community type and institutional resources function at the exosystem level, representing geographic and structural factors that influence educational experiences through resources allocation, community partnerships, and institutional capacity. Recent hierarchical analyses also highlight how district-level factors, such as poverty rates and racial composition, further shape these school-level effects (Ghimire & Regmi, 2020).

2.3.1. School socioeconomic composition

School socioeconomic composition represents one of the most influential mesosystem-level predictors of academic achievement across subject domains, reflecting how the aggregation of family backgrounds creates collective educational environments. Research consistently demonstrates that the aggregate socioeconomic profile of a school's student population accounts for substantial variance in achievement beyond individual SES factors (Perry & McConney, 2010). This school level contextual effect shapes the educational experiences of all students within that environment, regardless of their personal socioeconomic background.

PISA data analyses reveal that school socioeconomic composition often demonstrates stronger associations with achievement outcomes than individual SES (Liu et al., 2020), highlighting the powerful role of peer influences and institutional resources. Multilevel studies have shown that school-level factors typically account for 20–30% of total variance in mathematics performance, while the proportion varies for reading and science domains (Wang et al., 2023). This systematic review identified “school SES composition tended to be positively associated with math achievement” across 32 studies using PISA data (Wang et al., 2023, p. 17), although relationship strength varied significantly across different educational systems.

Multiple mechanisms mediate this relationship, including differential access to quality teachers, instructional resources, positive peer influences, and academically oriented school climates. School serving predominantly disadvantaged populations often struggle to recruit and retain highly qualified teachers, maintain adequate facilities, and establish rigorous academic expectations. Ghimire and Regmi (2020) observed comparable patterns in civics education, finding that a higher percentage of economically disadvantaged students, measured through free or reduced-price lunch eligibility, significantly correlated with lower proficiency rates on standardized civics assessments, reinforcing the systemic challenges posed by concentrated socioeconomic disadvantages.

2.3.2. *School community context*

The geographical and community setting of schools constitutes another critical exosystem -level dimension that shapes student achievement through resource availability, community partnerships, and institutional infrastructure. Research examining urban-rural performance disparities has generated inconsistent findings, suggesting complex interactions with factors such as socioeconomic composition and resource distribution rather than simple location effects (Drescher et al., 2022).

In their systematic review, Wang et al. (2023) identified 19 studies analyzing school location effects on mathematics achievement using PISA data. Eight studies reported positive associations between urban settings and mathematics performance, particularly in Australia, Spain, and the United States. However, other cross-national analyses revealed negative associations or highly context-dependent patterns varying by country. This variability suggests that community context effects are mediated by national educational policies, resource allocation systems, and other structural factors beyond simple urbanicity.

Community context frequently interacts with socioeconomic conditions in shaping educational outcomes. Drescher et al. (2022) demonstrated that while overall rural-urban achievement differences were modest, disparities became more pronounced for specific socioeconomic and racial-ethnic subgroups in certain geographic regions. Interestingly, the relationship between individual socioeconomic status and achievement appeared weaker in some rural contexts compared to urban settings, highlighting the complex interplay between location and other factors. Ghimire and Regmi (2020) provided additional evidence for these interaction effects, showing that district-level poverty rates and racial demographics significantly moderate school-level SES effects on standardized achievement, demonstrating how broader community characteristics can either amplify or buffer school-based disparities.

The availability of community resources, including cultural institutions, libraries, and collaborative relationships between schools and local organizations, creates differential opportunities for students across geographic contexts. These community assets vary systematically across urban, suburban, and rural environments, contributing to divergent achievement patterns across subject domains (Sulak, 2016).

2.4. *Cross-level interactions in educational outcomes*

Cross-level interactions between student and school characteristics represent a critical focus when examining educational outcomes through an ecological lens, as they reveal how factors operating at different system levels combine to shape achievement. These interactions embody Bronfenbrenner’s emphasis on the dynamic interplay between microsystem characteristics (individual student factors) and mesosystem/exosystem context (school environments). In educational contexts, school socioeconomic composition and community type can amplify or mitigate the effects of individual characteristics on achievement across subject domains (Perry & McConney, 2010). Recent PISA-based research has increasingly examined these specific moderating relationships, revealing how school contexts shape the influence of student demographic factors differently across reading, mathematics, and science (Wang et al., 2023).

The interaction between individual socioeconomic status and school socioeconomic composition represents one of the most extensively studied cross-level relationships. Research demonstrates that achievement gaps between high-SES and low-SES students often vary depending on school context. In some cases, these gaps are smaller in schools with higher average SES, suggesting compensatory effects for disadvantaged students (Xie & Ma, 2019), while other studies find that privileged students derive greater benefits from advantaged school environments, potentially widening disparities (Chiu & Chow, 2015). The mechanisms driving these interactions include peer effects, resource allocation patterns, instructional quality differences, and academic climate variations. Ghimire and Regmi (2020) found that district-level poverty rates significantly moderated school-level SES effects on achievement, demonstrating how broader contextual factors can alter the relationship between individual characteristics and outcomes.

Immigration status and language background interactions with school-level factors represent another important cross-level dynamic. Research indicates that school characteristics moderate achievement gaps between immigrant and native students and between students who speak the test language at home versus those who do not. Schools with higher proportion of immigrant students may develop more specialized support structures, potentially mitigating achievement gaps (OECD, 2018). However, high concentration of immigrant students in resource-constrained schools can exacerbate challenges related to language acquisition and academic integration (Borgonovi & Ferrara, 2020). Studies using PISA data have shown that language of instruction plays pivotal role in these interactions, with negative associations between non-test-language home environments and achievement being stronger in schools with linguistic support services.

The interaction between home and school technological resources has become increasingly relevant, particularly following pandemic-related educational disruptions. Research suggests that school ICT infrastructure and pedagogical technology integration can either amplify or mitigate the effects of home digital resources on achievement (Chiu, 2020). Students with limited technological access at home may particularly benefit from robust school-based digital resources, potentially reducing digital divides in educational outcomes. Conversely, limitations in both home and school technological environments can create compounded disadvantages. Ghimire and Mokhtari (2025) identified threshold effects in technology access, where basic resources supported achievement, but excessive device availability showed diminishing or negative returns, suggesting that optimal cross-level technology interactions involve thoughtful integration rather than mere abundance.

Teacher quality and instructional approaches also moderate the link between student characteristics and achievement (Kyriakides et al., 2020), with effects often manifesting differently across reading, mathematics, and science (Martinez-Abad et al., 2020). The COVID−19 pandemic likely altered these dynamics as teacher-student relationships adapted to remote and hybrid learning contexts.

3. Method

3.1. *Participants and sampling*

The sample for this study was drawn from the U.S. participants in the Programme for International Student Assessment (PISA) 2022, administered by the Organization for Economic Cooperation and Development (OECD). PISA is conducted every three years and assesses 15-year-old students' proficiency in reading, mathematics, and science, with each cycle focusing on one domain as the major area of assessment. The 2022 cycle featured mathematics as the major domain, with reading and science as minor domains. The assessment consisted of a two-hour cognitive test and background questionnaires completed by students and school principals.

PISA employs a two-stage stratified sampling design to ensure adequate representation of the target population. In the first stage, schools are selected with probability proportional to size; in the second stage, eligible students within each sampled school are randomly selected. The analytic sample for the present study consisted of 4552 students nested within 154 schools in the United States, representing the national population of approximately 3.3 million 15-year-old students.

3.2. *Variables and measures*

3.2.1. *Achievement scores*

Achievement in reading, mathematics, and science was measured using the ten plausible values (PV1-PV10) provided by PISA for each domain. Plausible values represent draws from the posterior distribution of each student's latent proficiency and account for measurement error inherent to the adaptive-testing framework (Mislevy, 1991; OECD, 2023). Correct analysis of plausible values requires fitting the substantive model separately on each PV and combining estimates using Rubin's (1987) rules, rather than averaging plausible values to produce a single composite score. We therefore treated each PV as a completed imputation of the latent achievement construct for every model reported below.

The analysis applies final student sampling weights (W_FSTUWT) at the student level when fitting each mixed-effects model. In PISA's two-stage sampling design, W_FSTUWT is the product of the school based weight (reflecting school selection probability) and a student base weight (reflecting within-school selection probability), adjusted for nonresponse; it therefore already incorporates both stages of selection probabilities into a composite weight (NCES, 2019; OECD, 2023). Consistent with recent methodological work, this study acknowledges that specifying weights separately at each level may improve variance-component estimation in multilevel models (Atasever et al., 2025; Mang et al., 2021). Because {lme4} does not natively support weights at multiple levels, the variance components reported in this study (ICCs, random effect variances) should be interpreted with appropriate caution, whereas fixed-effect estimates are expected to be robust. To stabilize variance-component estimation within each fit, W_FSTUWT was normalized to have a within-sample mean of 1 before being passed to *lmer()*as case weights; this rescaling preserves relative weighting while preventing the raw PISA weights (mean ≈ 804) from inflating the estimated residual variance.

3.2.2. *Student-Level Variables*

- *Gender*: Students' self-reported binary variable (female, male)
- *Immigration Background (IMMIG)*: Categorical variable classified into native, first-generation, and second-generation immigrant status.
- *Parental Education (PAREDINT)*: Highest education level of either parent, reported in years of formal schooling. Parental education was selected rather than the composite ESCS index for three reasons: (1) it provides more interpretable coefficients representing each additional year of schooling; (2) it facilitates cleaner estimation of compositional effects when school means are computed; and (3) it avoids collinearity with ICT resources, which are examined separately but overlap with the home possessions components of ESCS.
- *Home ICT Resources (ICTRES)*: Standardized index (M = 0, SD = 1 in OECD countries) measuring the availability of digital technology resources at home, derived from student questionnaire items assessing access to computers, tablets, internet connectivity, and educational software. Higher values indicate greater availability of home ICT resources.
- *Language Spoken at Home (LANGN)*: Primary home language, categorized specifically for the U.S. sample as English, Spanish, or neither (another language).

3.2.3. *School-level variables*

- *School Community Type*: Categorized by population size (fewer than 15,000 residents, 15,001 – 100,000 residents, 100,001 – 1 million residents, over 1 million residents).
- *Percentage of Students with Special Learning Needs*: Proportion of students with special learning needs as reported by principals.
- *Percentage of Socioeconomically Disadvantaged Students*: Principals' estimation of socioeconomically disadvantaged students in respective school.
- *School ICT Availability (ICTSCH)*: Standardized index (M = 0, SD = 1 in OECD countries) measuring institutional digital infrastructure, derived from school principal questionnaire items about availability of computers, internet bandwidth, digital learning platforms, and technical support. Higher values indicate more extensive school ICT infrastructure.

3.3. *Missing data analysis and handling*

Prior to the main analysis, a comprehensive examination of missing data patterns was conducted. The analysis revealed varying levels of missingness across variables. Student-level variables showed moderate to low missingness: immigration status (8.7%), school community type (5.1%), parental education (4.9%), ICT resources (4%), and gender (0.1%).

Chi-square tests revealed significant relationships between missing parental education data and both gender ($x^2 = 6.19$, $p = 0.013$) and immigration status ($x^2 = 21.62$, $p < 0.001$), indicating that they were not missing completely at random. Furthermore, *t*-test comparing students with complete versus incomplete data showed significant differences in academic performance across all domains (reading: $t = -7.57$, $p < .001$; mathematics: $t = -6.78$, $p < .001$; science: $t = -7.22$, $p < .001$), with students having complete data consistently outperforming those with missing values.

Given the systematic patterns in missingness identified through preliminary analyses (Little, 1988), multiple imputation was implemented with the {mice} package in R (van Buuren & Groothuis-Oudshoorn, 2011). Ten imputed datasets were generated with five iterations each, using a consistent random seed (2025) to ensure reproducibility. The imputation model incorporated both student- and school-level variables to preserve the hierarchical structure of the data. For each variable, mice selected an appropriate univariate method based on the variable's measurement level, with sample (draw from observed values) used for categorical variables with many levels to maintain their distributional properties. Density-plot comparisons between observed and imputed values confirmed adequate imputation quality. The ten imputed datasets were used in combination with the ten plausible values per outcome, yielding 100 fits per model specification per subject.

3.4. *Analytical approach*

To account for the nested structure of students within schools, hierarchical linear modeling (HLM) was employed using the {lme4} package in R (Bates et al., 2015). This approach addresses the non-independence of observations and allows for the partitioning of variance between school and student levels, a critical consideration when analyzing PISA data (Rutkowski et al., 2010). The analysis proceeded through the following steps:

First, unconditional (null) models were estimated separately for reading, mathematics, and science achievement to calculate intraclass correlation coefficients (ICCs), representing the proportion of variance in achievement attributable to between-school differences.

Next, a series of increasingly complex models was estimated for each achievement domain:

- *Model 1 (Null/Unconditional)*: Contains only the school random intercept. Used to establish the null ICC and baseline variance partitioning.
- *Model 2 (Student-Level).* Adds student-level predictors: gender, home language, immigration status, parental education, and home ICT resources.
- *Model 3 (Full Model)*: Adds school-level predictors: school community type, percentage of socioeconomically disadvantaged students, percentage of students with special learning needs, and school ICT availability.
- *Model 4 (Contextual Effects)*: Replaces raw individual-level parental education and home ICT resources with their group-mean-centered counterparts and adds corresponding school-mean aggregates. Group-mean centering (subtracting each school's mean from its students' scores) does not only separate the within-school association from the between-school (compositional) association but also provides more interpretable coefficients with reduced multicollinearity (Enders & Tofighi, 2007; Lüdtke et al., 2008; Lüdtke et al., 2009). This specification allows estimation of both individual- and school-level effects simultaneously.
- *Model 5 (Cross-Level Interactions)*: Extends Model 4 by adding interactions between student- level demographics (home language, immigration status, gender) and school-level contextual factors (percentage of disadvantaged students, community type). These interactions test whether school context moderates individual-level associations with achievement.
- *Model 6 (Random Slopes Models)*: Allows the effects of key individual-level predictors (gender, home language, and parental education) to vary across different schools. This model quantifies between-school heterogeneity in demographic associations that a fixed-slope specification would otherwise conceal.

The general form of the full two-level model with cross-level interactions was:

Level 1 (Student):

$$\begin{aligned} Y_{i}j = {} & \beta_{0}j + \beta_{1}j(\llbracket Gender \rrbracket_{i}j) + \beta_{2}j(\llbracket HomeLanguage \rrbracket_{i}j) \\ & + \beta_{3}j(\llbracket ImmigrationStatus \rrbracket_{i}j) + \beta_{4}j(\llbracket ParentalEducation \rrbracket_{i}j) \\ & + \beta_{5}j(\llbracket ICTResources \rrbracket_{i}j) + r_{i}j \end{aligned} \tag{1}$$

Level 2 (School):

$$\begin{aligned} \beta_{0}j = {} & \gamma_{0}0 + \gamma_{0}1(\llbracket SchoolCommunityType \rrbracket_{j}) \\ & + \gamma_{0}2(\llbracket \%DifferentHeritageLanguage \rrbracket_{j}) \\ & + \gamma_{0}3(\llbracket \%ParentsImmigrated \rrbracket_{j}) + \gamma_{0}4(\llbracket SchoolICTResources \rrbracket_{j}) + u_{0}j \end{aligned} \tag{2}$$

$$\beta_{2}j = \llbracket \gamma_{2}0 + \gamma \rrbracket_{2}1(\llbracket \%DifferentHeritageLanguage \rrbracket_{j}) + u_{2}j \tag{3}$$

$$\beta_{3}j = \gamma_{3}0 + \gamma_{3}1(\llbracket \%ParentsImmigrated \rrbracket_{j}) + u_{3}j \tag{4}$$

Where,

- ○ Y_{ij} = Achievement score for student *i* in school *j*.
- ○ β_{0j} = Adjusted mean achievement scores for school *i*.
- ○ β_{1j} through β_{5j} = Coefficients for student-level predictors (Gender, Home Language, Immigration Status, Parental Education, ICT Resources).
- ○ r_{ij} = Student-level residual error.
- ○ γ_{00} = Grand mean achievement across all schools.
- ○ γ_{0j} through γ_{5j} = Effects of school-level predictors (School Community Type, % Different Heritage Language, % Parents Immigrated, School ICT Resources) on school mean achievement.
- ○ γ_{20}, γ_{30} = Average effects of home language and immigration status across schools, respectively.
- ○ γ_{21}, γ_{31} = Moderating effects of school-level compositional factors (% Different Heritage Language, %Parents Immigrated) on individual-level relationships.
- ○ u_{0j}, u_{2j}, u_{3j} = School-level random errors terms associated with the intercept and slopes for Home Language and Immigration Status, respectively.

Each of the six model specifications was fit separately on every combination of the 10 plausible values and the 10 imputed datasets, yielding 100 fits per model per subject (600 fits per subject total, 1800 fits across the three achievement domains). All models were estimated with lmer() in the {lme4} package (Bates et al., 2015). Fixed-effect estimates, standard errors, and variance components were pooled across the 100 fits per model using Rubin's (1987) rules as implemented in mitml::testEstimates() (Grund et al., 2021), which propagates both within-fit sampling variance and between-fit imputation variance. This treatment of PVs as imputations of the latent achievement construct is the statistically correct approach to multiply-imputed proficiency scores (OECD, 2023; von Davier et al., 2009) and avoids the downward bias in standard errors that results from averaging PVs into a single composite score before estimation.

Marginal R^2 was computed for each fit using MuMIn::r.squaredGLMM() (Nakagawa & Schielzeth, 2013) and averaged across fits for presentation. The ICC for each model was derived from the pooled τ_{00} and σ^2 estimates. Fixed-effect estimates are expected to be robust; variance-component estimates should be interpreted with appropriate caution for the reasons noted above (single-level weighting in {lme4}). The large student sample (n = 4552) and substantial number of Level−2 units (n = 154 schools) provide adequate power for detecting meaningful fixed effects.

3.5. *Transparency and openness*

This study follows the Journal Article Reporting Standards (JARS; Appelbaum et al., 2018). The data used in this study are publicly available through the OECD PISA database (https://www.oecd.org/pisa/data/2022database/). All analyses were conducted using R version 4.5.3 (R Core Team, 2026), with specialized packages for handling complex survey data (survey; Lumley, 2010), multiple imputation (mice, mitools; van Buuren & Groothuis-Oudshoorn, 2011), and multilevel modeling (lme4, lmerTest; Bates et al., 2015). The multilevel models were built sequentially as described, and all parameters and exclusion criteria are reported in the text. The complete analysis code, data processing scripts, and supplementary materials are publicly available on the Open Science Framework at https://osf.io/fdbgj/.

4. Results

4.1. *Descriptive statistics*

Table 1 presents the descriptive statistics for the sample of 4552 students from 154 schools. The sample was balanced by gender (49.1% female, 50.8% male), with missing data accounting for 0.12%. The majority of students were native born (69.97%), with second-generation immigrants comprising 16.65% and first-generation immigrants 4.63% with 8.75% having missing immigration status data. The mean year of parental education was 14.06, indicating an average educational attainment beyond high school.

School characteristics varied by community types, with the largest proportion of schools located in communities of 15,000–100,000 residents (29.22%), and indicated that, on average, 46.17% of students were

Table 1
Descriptive Statistics for Student and School Characteristics and Achievement Scores.

Panel A: Student Level Characteristics								
	N	%	Reading (SD)	Math(SD)	Science(SD)			
Gender								
Female	2235	49.10	514.6 (100.8)	458.1(84.1)	495.8(97.5)			
Male	2312	50.79	493.0 (109.9)	471.4(96.9)	502.7(109.8)			
Immigration Status								
Native	3185	69.97	511.1 (103.7)	470(89.6)	507.6(101.4)			
First-Generation	211	4.64	466.3 (118.9)	441.1(97.7)	460.8(109.7)			
Second-Generation	758	16.65	505.5 (107.4)	465.7(94.9)	496.6(108.2)			
Home Language								
English	3494	76.76	513.8 (103.8)	472.2(90.4)	509.9(101.8)			
Spanish	650	14.28	451.1 (102.8)	418.7(79.9)	441.5(98.7)			
Another Language	200	4.39	513.3 (117.1)	485.2(104.6)	512.3(112.4)			
	Mean	SD	Min	Max				
Parental Education	14.06	2.67	3	16				
ICT Resources at Home	0.21	0.96	−5.03	5.17				
Panel B: School Level Characteristics								
School Community Type								
Community < 15k	36	23.38	495.8(99.0)	461.8(82.6)	495.7(97.0)			
Community 15–100k	45	29.22	514.1 (105.3)	474.7(91.8)	510.4(105.2)			
Community 100k – 1 M	40	25.97	502.8 (112.4)	459.5(96.9)	496.5(109.2)			
Community > 1 M	25	16.23	505.7 (106.7)	462.8(90.8)	493.4(103.9)			
	Mean	SD	Min	Max				
ICT Available at School	0.2035	0.5251						
% Socioeconomically Disadvantaged Students	46.17	29.16	0.00	100				
% Special Learning Needs	14.96	10.79	1.00	100				
Panel C: Achievement Scores								
	Mean	SD	Min	Max				
Reading	503.94	106.06	107.96	798.18				
Mathematics	464.89	91.05	224.89	785.65				
Science	499.41	103.94	187.92	786.47				
Panel D: Correlation Matrix for Key Variables								
	ICTSCH	ICTRES	PAREDINT	% Special Learning Needs	% Socioeconomically Disadvantaged Students	Math	Reading	Science
ICTSCH	1.000							
ICTRES	0.025	1.000						
PAREDINT	0.027	0.208	1.000					
% Special Learning Needs	−0.003	0.006	0.018	1.000				
% Socioeconomically Disadvantaged Students	−0.070	−0.171	−0.256	0.079	1.000			
Math	0.067	0.149	0.209	0.026	−0.340	1.000		
Reading	0.074	0.135	0.174	0.021	−0.291	0.888	1.000	
Science	0.081	0.136	0.196	0.032	−0.327	0.932	0.919	1.000

Note. N represents the number of cases and percentages are based on valid responses. Achievement scores are presented as mean (SD) of 10 plausible values in this table (standard OECD convention for descriptive reporting). All inferential models use proper PV-as-imputation pooling via Rubin's rules. ICTSCH = ICT available at school; ICTRES = ICT resources at home; PAREDINT = parental education; "%Special Learning Needs" indicates the percentage of students with special learning needs; "% Socioeconomically Disadvantaged Students" indicates the percentage of students classified as socioeconomically disadvantaged. Correlations in Panel D are Pearson's r coefficients computed with pairwise deletion of missing data.

socioeconomically disadvantaged, while 14.96% had special learning needs. The weighted average achievement scores, computed as student-level means across 10 plausible values per domain (standard OECD descriptive convention) and then averaged across the analytic sample, were: reading (M = 503.94, SD = 106.06), science (M = 499.41, SD = 103.94), and mathematics (M = 464.89, SD = 91.05). These means are based on the complete-case sample (n = 4552) after multiple imputation of missing covariates and differ slightly from OECD-reported U.S. means of 504 (reading), 465 (mathematics), and 499 (science) for the full U.S. sample. The small deviations reflect differences in sample composition after excluding cases with missing school-level data and applying this study's analytic sample restrictions. The sample in this study maintains representativeness of the U.S. 15-year-old population through application of student sampling weights (W_FSTUWT).

One data-quality disclosure is warranted for school ICT availability (ICTSCH). In the U.S. PISA 2022 sample, 112 of the 135 schools with valid ICTSCH data (83%) clustered at the index ceiling of 0.4062, with the remaining 23 schools distributed across lower values (range −2.85 to −0.08) and no observations above the ceiling. This clustering suggests that, in the U.S. context, ICTSCH functions less as a finely graded scale and more as an indicator of specific school-reported ICT deficiencies. ICTSCH coefficients reported below should therefore be interpreted as capturing the contrast between the modal well-resourced and the subset of lower-scoring schools, rather than a smooth dose-response relationship.

Pearson's correlation coefficients indicated strong intercorrelations among these domains, ranging from 0.888 to 0.932. Moreover, parental education was moderately positively associated with achievement (r = 0.180–0.212), whereas the percentage of socioeconomically disadvantaged students showed negative associations (r = −0.291 to −0.340). Detailed descriptive statistics are presented in Table 1.

RQ1. : Variance Decomposition Between Student and School Levels

To address the first research question regarding the proportion of variance attributable to student versus school factors, the analysis began with unconditional (null) models for each subject domain. The pooled intraclass correlation coefficients (ICCs) indicated substantial between-school variance, with mathematics showing the highest proportion of school-level variance (ICC = 0.245), followed by science (ICC = 0.224) and reading (ICC = 0.189). These findings indicate that approximately 19–25% of the total variance in achievement scores is attributable to between-schools differences after pooling across the 10 PVs and imputations, with the remaining variance occurring at the student level.

As shown in Tables 2–4 (Model 1), the null models established baseline variance partitioning across three domains. The pooled school-level variance component was largest for science ($\tau_{00} = 2652.19$), followed by reading ($\tau_{00} = 2358.70$) and mathematics ($\tau_{00} = 2212.63$). The progression from Model 1 through Model 6 demonstrated systematic reductions in unexplained school-level variance as predictors were added, with the exception of Model 6 (Random Slopes), in which τ_{00} rebounds modestly as variance shifts from the random intercept into the newly estimated random slopes. Fig. 1 illustrates variance components across model specifications, showing the proportion of total variance explained by fixed effects (R^2, displayed as bars) and the intraclass correlation coefficient (ICC, displayed as circles connected by dashed lines). The ICC represents the proportion of residual variance occurring between schools after accounting for model predictors; its decline from Model 1 to Model 5 reflects the addition of student- and school-level predictors, whereas the slight rebound at Model 6 reflects the redistribution of variance into random-slope components rather than a failure of model fit.

RQ2. : Individual Background Factors and Domain-Specific Achievement

The second research question examined how individual background factors relate to performance across subject domains. Findings from Models 2 and 3 in Tables 2–4 and summarized visually in Fig. 2, revealed

Table 2

Multilevel Models for Reading Achievement.

Fixed Effects	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
	Null Model	Student-Level Model	Full Model	Contextual Model	Cross-Level Interactions	Random Slopes
Intercept	502.23(4.29)***	431.67(11.22)***	451.61(14.30)***	345.00(58.35)***	339.14(58.80)***	367.98(57.53)****
Student-Level Variables						
Gender (Male)	-	−22.00(3.07)***	−22.11(3.09)***	−22.11(3.08)***	−26.39(6.51)***	−22.83(3.30)***
Home Language (English)	-	46.85(6.26)***	46.68(6.31)***	45.60(6.28)***	54.37(13.53)***	45.01(7.07)***
Home Language (Other)	-	31.76(9.05)***	31.50(9.04)***	30.63(9.03)**	48.75(17.51)**	26.43(10.14)**
Immigration Status (2nd Gen)	-	18.22(5.69)**	18.95(5.69)**	19.72(5.68)**	20.33(11.35)~	19.56(5.65)**
Immigration Status (1st Gen)	-	−12.13(9.11)	−11.61(9.14)	−11.70(9.11)	−20.68(17.29)	−13.06(9.20)
Parental Education (Within)	-	-	-	2.35(0.71)**	2.36(0.72)**	2.38(0.79)**
Parental Education	-	2.77(0.71)***	2.64(0.71)***	-	-	-
ICT Resources (Within)	-	-	-	3.60(1.78)*	3.50(1.79)~	3.64(1.79)*
ICT Resources	-	4.46(1.77)*	4.27(1.78)*		-	-
School-Level Variables						
School ICT Resources	-	-	3.68(2.04)~	3.60(2.03)~	3.69(2.04)~	3.39(2.03)~
Community (15–100k)	-	-	14.43(9.04)	15.40(8.80)~	9.93(9.88)	15.18(8.61)~
Community (100k-1M)	-	-	10.56(9.13)	11.43(8.94)	11.47(9.80)	9.96(8.74)
Community (>1 M)	-	-	10.09(10.94)	11.60(10.72)	7.74(12.32)	11.35(10.47)
% Socioeconomic Disadvantaged	-	-	−0.62(0.12)***	−0.41(0.13)**	−0.28(0.24)	−0.42(0.13)**
% Special Learning Needs	-	-	0.07(0.31)	0.04(0.30)	0.04(0.30)	0.06(0.30)
Compositional Effects						
School Mean Parental Education	-	-	-	9.02(4.03)*	9.04(4.00)*	7.57(3.98)~
School Mean ICT Resources	-	-	-	42.36(14.76)**	41.87(14.64)**	38.60(14.50)**
Interactional Effects						
English X School SES	-	-	-	-	−0.16(0.21)	-
Other Language X School SES	-	-	-	-	−0.37(0.33)	-
2nd Gen X School SES	-	-	-	-	−0.02(0.20)	-
1st Gen X School SES	-	-	-	-	0.18(0.29)	-
Male X Community (15–100k)	-	-	-	-	10.95(8.92)	-
Male X Community (100k-1M)	-	-	-	-	−0.20(8.42)	-
Male X Community (>1 M)	-	-	-	-	7.17(10.98)	-
Random Effects						
τ_{00} (School Variance)	2358.70	1925.96	1325.29	1206.30	1186.25	1804.19
σ^2 (Residual Variance)	10113.25	9724.53	9747.80	9730.83	9734.95	9506.80
Model Fit						
Marginal $\underline{R}^2$	0.000	0.046	0.093	0.129	0.117	0.108
ICC	0.189	0.165	0.120	0.110	0.109	0.159
School *N*	154	154	154	154	154	154
Student *N*	4552	4552	4552	4552	4552	452

Note. Values are unstandardized coefficients with standard errors in parentheses, pooled across 10 plausible values X 10 imputation via Rubin's rules. ICC = Intraclass Correlation Coefficient, τ_{00} = school-level variance; σ^2 = residual variance; Marginal $\underline{R}^2$ = proportion variance explained by fixed effects, averaged across fits. Reference categories: Gender (Female), Home Language (Spanish), Immigration Status (Native), Community Type (<15k). Model 4 uses group-mean-centered individual predictors with school-mean aggregates. Model 6 allows PAREDINT, gender, and LANGN slopes to vary by school. ~$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$.

Table 3
Multilevel Models for Mathematics Achievement.

Fixed Effects	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
	Null Model	Student-Level Model	Full Model	Contextual Model	Cross-Level Interactions	Random Slopes
Intercept	463.36(4.04)***	380.94(9.31)***	402.57(12.53)***	280.09(53.42)***	264.98(53.54)***	318.15(51.10)***
Student-Level Variables						
Gender (Male)	-	12.62(2.55)***	12.55(2.56)***	12.56(2.56)***	10.07(5.49)∼	11.82(2.66)***
Home Language (English)	-	32.16(5.22)***	32.08(5.26)***	31.14(5.22)***	48.25(10.55)***	31.87(6.02)***
Home Language (Other)	-	30.38(7.66)***	30.05(7.66)***	29.26(7.63)***	49.14(13.92)***	27.32(8.71)**
Immigration Status (2nd Gen)	-	14.95(4.54)**	15.65(4.53)**	16.42(4.52)***	19.32(9.34)*	15.52(4.58)**
Immigration Status (1st Gen)	-	−3.33(7.43)	−2.73(7.47)	−2.78(7.44)	−10.83(14.74)	−4.74(7.54)
Parental Education (Within)	-	-	-	2.90(0.59)***	2.88(0.59)***	2.97(0.68)***
Parental Education	-	3.22(0.58)***	3.14(0.58)***	-	-	-
ICT Resources (Within)	-	-	-	3.95(1.52)**	3.91(1.52)*	3.80(1.51)*
ICT Resources	-	4.60(1.50)**	4.50(0.58)***	-	-	-
School-Level Variables						
School ICT Resources	-	-	2.59(1.63)	2.53(1.62)	2.65(1.63)	2.44(1.62)
Community (15–100k)	-	-	9.14(8.76)	10.39(8.43)	7.24(9.47)	8.99(8.16)
Community (100k-1M)	-	-	2.92(8.07)	4.04(7.82)	4.87(8.61)	4.18(7.62)
Community (>1 M)	-	-	−0.31(10.07)	1.61(9.71)	−2.03(10.88)	1.40(9.38)
% Socioeconomic Disadvantaged	-	-	−0.57(0.11)***	−0.37(0.12)**	−0.11(0.20)	−0.37(0.12)**
% Special Learning Needs	-	-	0.12(0.27)	0.09(0.26)	0.10(0.26)	0.10(0.26)
Compositional Effects						
School Mean Parental Education	-	-	-	10.65(3.66)**	10.76(3.62)**	8.07(3.51)*
School Mean ICT Resources	-	-	-	42.95(13.22)**	42.31(13.09)**	37.04(12.86)**
Interactional Effects						
English X School SES	-	-	-	-	−0.31(0.17)∼	-
Other Language X School SES	-	-	-	-	−0.38(0.26)	-
2nd Gen X School SES	-	-	-	-	−0.06(0.16)	-
1st Gen X School SES	-	-	-	-	0.16(0.24)	-
Male X Community (15–100k)	-	-	-	-	6.52(7.50)	-
Male X Community (100k-1M)	-	-	-	-	−1.70(7.50)	-
Male X Community (>1 M)	-	-	-	-	6.74(9.01)	-
Random Effects						
τ_{00} (School Variance)	2212.63	1813.35	1246.40	1092.23	1062.87	1006.07
σ^2 (Residual Variance)	6809.19	6592.17	6610.91	6596.59	6595.61	6385.51
Model Fit						
Marginal $\underline{R}^2$	0.000	0.037	0.086	0.138	0.140	0.117
ICC	0.245	0.216	0.159	0.142	0.139	0.136
School *N*	154	154	154	154	154	154
Student *N*	4552	4552	4552	4552	4552	452

Note. Values are unstandardized coefficients with standard errors in parentheses, pooled across 10 plausible values X 10 imputation via Rubin's rules. ICC = Intraclass Correlation Coefficient, τ_{00} = school-level variance; σ^2 = residual variance; Marginal $\underline{R}^2$ = proportion variance explained by fixed effects, averaged across fits. Reference categories: Gender (Female), Home Language (Spanish), Immigration Status (Native), Community Type (<15k). Model 4 uses group-mean-centered individual predictors with school-mean aggregates. Model 6 allows PAREDINT, gender, and LANGN slopes to vary by school. ∼$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$.

both consistent and domain-specific patterns.

Gender demonstrated the most pronounced domain-specific patterns. In Model 3 (Full Model), males underperformed in reading by −22.11 points (SE = 3.09, $p < .001$) but outperformed females in mathematics by 12.55 points (SE = 2.56, $p < .001$). In science, the male advantage was small and of borderline significance (6.16 points, SE = 3.20, $p = .055$). These gender coefficients remained stable across Models 2 through 4 and 6, with only Model 5 (CLI) showing slightly altered estimates because gender is involved in interactions with community type.

Home language showed consistent positive associations across all domains. English-speaking students scored substantially higher than Spanish-speaking peers (the reference category) in every model. In Model 3, English speakers demonstrated advantages of 46.68 points in reading (SE = 6.31), 32.08 points in mathematics (SE = 5.26), and 45.56 points in science (SE = 6.04), all $p < .001$. Students whose home language was neither English nor Spanish demonstrated intermediate but still substantial advantages over Spanish speakers: 31.50 in reading, 30.05 in mathematics, and 35.37 points in science (all $p < .001$).

Immigration status showed nuanced and partly attenuated patterns. Second-generation immigrant students consistently outperformed native students across all three domains in Model 3: 18.95 points in reading (SE = 5.69, $p < .001$), 15.65 points in mathematics (SE = 4.53, $p < .001$), and 14.74 points in science (SE = 5.29, $p < .001$). First-generation immigrant students, however, showed point estimates in the negative direction but with confidence intervals that included zero once measurement uncertainty in the plausible values was properly propagated. The wider confidence intervals for first-generation students partly reflect their smaller subsample (n = 211) and properly propagated PV uncertainty.

Parental education showed consistent positive associations with achievement across all three domains. In Model 4, where parental education was group-mean centered to isolate the within-school association, each additional year was associated with 2.35 points in reading (SE = 0.71, $p < .001$), 2.90 points in mathematics (SE = 0.59, $p < .001$), and 2.71 points in science (SE = 0.71, $p < .01$). These within-school coefficients were slightly smaller than uncentered estimates in Model 2 and 3 (roughly 2.50–3.20), reflecting the introduction of school-mean parental education as a separate predictor.

Home ICT resources also showed positive associations across domains. In Model 3 (Full Model), each standard-deviation unit of ICTRES was associated with 4.27 points in reading (SE = 1.78, $p < .05$), 4.50 in mathematics (SE = 1.51, $p < .01$), and 3.93 points in science (SE = 1.95, $p < .01$). Similar to parental education, these individual-level coefficients decreased in Model 4 when school-mean ICT resources was added.

Adding school-level predictors in Model 3 reduced school-level variance from roughly 20–22–15–17 across domains and ICCs

Table 4
Multilevel Models for Science Achievement.

Fixed Effects	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
	Null Model	Student-Level Model	Full Model	Contextual Model	Cross-Level Interactions	Random Slopes
Intercept	498.00(4.48)***	410.48(10.64)***	432.25(14.58)***	321.10(61.79)***	306.97(62.01)***	340.54(59.48)***
Student-Level Variables						
Gender (Male)	-	6.20(3.19)~	6.16(3.20)~	6.17(3.20)~	3.46(6.44)	5.33(3.46)
Home Language (English)	-	45.69(6.00)***	45.56(6.04)***	44.58(6.04)***	60.83(12.76)***	44.82(6.98)***
Home Language (Other)	-	35.76(8.38)***	35.37(8.36)***	34.55(8.32)***	50.92(15.94)**	32.80(9.59)**
Immigration Status (2nd Gen)	-	13.93(5.31)**	14.74(5.29)**	15.46(5.30)**	17.98(10.93)	15.16(5.35)**
Immigration Status (1st Gen)	-	−12.66(8.63)	−11.91(8.64)	−11.98(8.61)	−21.21(16.76)	−12.15(8.75)
Parental Education (Within)	-	-	-	2.71(0.71)***	2.70(0.71)***	2.76(0.81)**
Parental Education	-	3.08(0.70)***	2.97(0.70)***	-	-	-
ICT Resources (Within)	-	-	-	3.31(1.97)~	3.26(1.97)~	3.16(1.96)
ICT Resources	-	4.08(1.94)*	3.93(1.95)*	-	-	-
School-Level Variables						
School ICT Resources	-	-	4.52(1.95)*	4.44(1.94)*	4.57(1.95)*	4.38(1.93)*
Community (15–100k)	-	-	12.28(9.25)	13.26(9.09)	9.64(10.06)	12.19(8.86)
Community (100k-1M)	-	-	7.56(9.27)	8.43(9.07)	9.15(9.81)	7.94(8.79)
Community (>1 M)	-	-	0.73(11.00)	2.25(10.81)	−0.92(12.02)	0.57(10.57)
% Socioeconomic Disadvantaged	-	-	−0.63(0.13)***	−0.42(0.14)**	−0.18(0.24)	−0.44(0.14)*
% Special Learning Needs	-	-	0.16(0.31)	0.12(0.30)	0.13(0.29)	0.15(0.29)
Compositional Effects						
School Mean Parental Education	-	-	-	9.63(4.20)*	9.73(4.16)*	8.44(4.06)*
School Mean ICT Resources	-	-	-	44.81(14.86)**	44.30(14.75)**	39.05(14.63)**
Interactional Effects						
English X School SES	-	-	-	-	−0.29(0.20)	-
Other Language X School SES	-	-	-	-	−0.29(0.30)	-
2nd Gen X School SES	-	-	-	-	−0.05(0.19)	-
1st Gen X School SES	-	-	-	-	0.18(0.29)	-
Male X Community (15–100k)	-	-	-	-	7.51(8.59)	-
Male X Community (100k-1M)	-	-	-	-	−1.55(8.19)	-
Male X Community (>1 M)	-	-	-	-	5.86(10.28)	-
Random Effects						
τ_{00} (School Variance)	2652.19	2171.44	1488.31	1359.20	1328.69	1841.62
σ^2 (Residual Variance)	9169.64	8874.80	8893.42	8875.24	8877.82	8577.02
Model Fit						
Marginal $\underline{R}^2$	0.000	0.039	0.085	0.126	0.128	0.116
ICC	0.224	0.197	0.143	0.133	0.130	0.177
School *N*	154	154	154	154	154	154
Student *N*	4552	4552	4552	4552	4552	452

Note. Values are unstandardized coefficients with standard errors in parentheses, pooled across 10 plausible values X 10 imputation via Rubin's rules. ICC = Intraclass Correlation Coefficient, τ_{00} = school-level variance; σ^2 = residual variance; Marginal $\underline{R}^2$ = proportion variance explained by fixed effects, averaged across fits. Reference categories: Gender (Female), Home Language (Spanish), Immigration Status (Native), Community Type (<15k). Model 4 uses group-mean-centered individual predictors with school-mean aggregates. Model 6 allows PAREDINT, gender, and LANGN slopes to vary by school. ~p < .10, *p < .05, **p < .01, ***p < .001.

from.18t to .24–.13-.17, indicating that measured school characteristics account for a meaningful share of between-school differences.

RQ3. : School-Level Factors and Compositional Effects

Model 3 (Full Model) introduced school-level variables to examine their associations with student achievement. School socioeconomic composition, measured as the percentage of socioeconomically disadvantaged students, emerged as the strongest school-level predictor across all domains. Each additional percentage point of disadvantaged students was associated with decreases of 0.62 points in reading (SE = 0.12), 0.57 points in mathematics (SE = 0.11), and 0.63 points in science (SE = 0.13), all $p < .001$, after controlling for individual background factors. The percentage of students with special learning needs showed minimal and non-significant associations across domains.

School ICT availability (ICTSCH) showed domain-specific associations. It was positively and significantly related to science achievement (4.52, SE = 1.95, $p < .05$), and marginally related to reading (3.68 points, SE = 2.04, $p = .071$). The mathematics coefficient was smaller and not significant (2.59 points, *SE* = 1.63, $p = .11$). These patterns should be interpreted in light of the ICTSCH ceiling effect noted in the Descriptive Statistics section: 83% of U.S. schools cluster at the index ceiling, so ICTSCH coefficients primarily reflect the contrast between the modal well-resourced schools and a smaller subset of lower-scoring schools. School community type showed modest and non-significant associations with achievement, with students in communities of 15,000−100,000 residents scoring 14.43, 9.14, and 12.28 points higher in reading, mathematics, and science, respectively, than those in communities with fewer than 15,000 residents (all $p > .05$).

4.2. *Compositional and contextual factors*

The contextual effects models (Model 4 in Tables 2–4) distinguished between within-school (individual) and between-school (compositional) associations by combining group-mean-centered individual predictors with school-mean aggregates. School-mean parental education was substantially more strongly associated with achievement than individual-level parental education: 9.02 points per year in reading (SE = 4.03, $p < .05$), 10.65 points in mathematics (SE = 3.66, $p < .01$), and 9.63 points in science (SE = 4.20, $p < .05$). These between-school associations were approximately 3.7–3.8 times larger than the corresponding individual-level coefficients (2.35, 2.90, and 2.71 points, respectively).

School-mean home ICT resources also showed substantially larger between-school than within-school associations: 42.36 points in reading (SE = 14.76, $p < .01$), 42.95 points in mathematics (SE = 13.22, $p < .01$), and 44.81 points in science (SE = 14.86, $p < .001$). Fig. 3 visualizes the within- vs. between-school contrasts for parental education, with 95% confidence intervals that make explicit the greater uncertainty

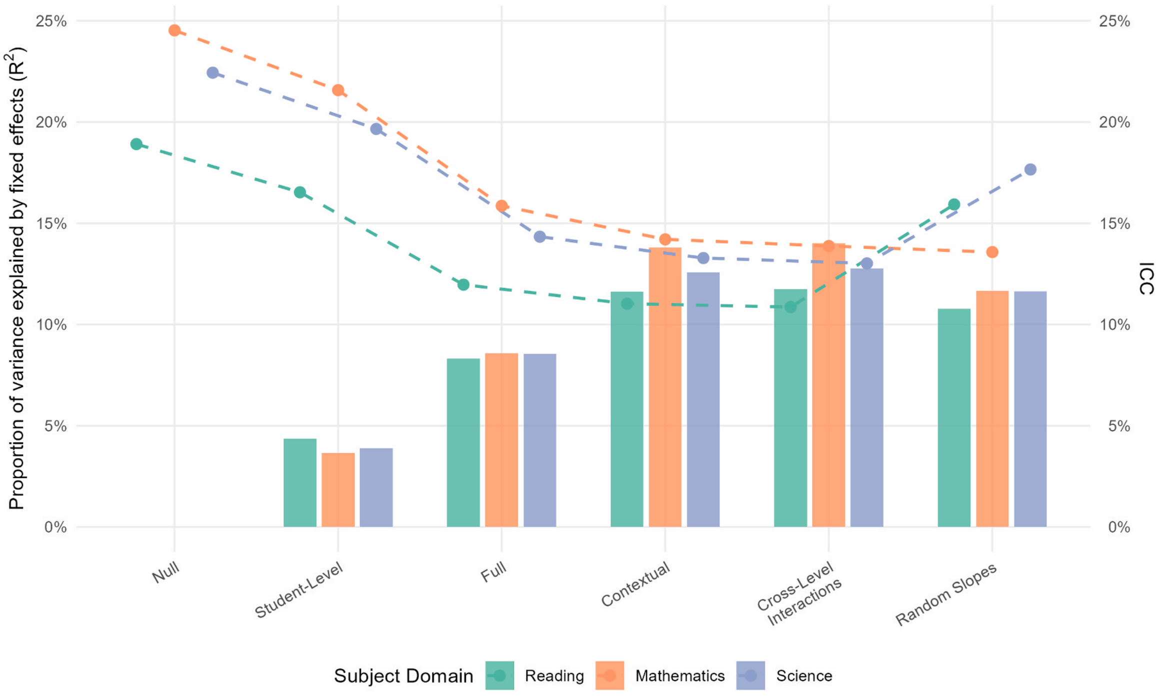

Fig. 1. Variance Decomposition Across Model Specifications. Note. Bars represent the proportion of total variance explained by fixed effects (R^2) at each model stage, averaged across the 100 fits per model. Circles connected by dashed lines show the intraclass correlation coefficient (ICC), indicating the proportion of residual variance occurring between schools. ICCs decline as school-level predictors are added (Models 2–5), demonstrating that measured school characteristics account for part of the between-school variance. The slight rebound at Model 6 (Random Slopes) reflects the redistribution of variance from the random intercept into random slopes. Model specifications: Null (unconditional); Student-Level (individual predictors only); Full (adds school-level predictors); Contextual (group-mean centered individual predictors with school-mean aggregates); Cross-Level Interactions (tests moderation effects); Random Slopes (allows gender, English vs. Spanish, and within-school parental education to vary by school).

of school-level aggregates given the smaller Level−2 N and the use of manifest rather than latent school means (Lüdtke et al., 2008; Lüdtke et al., 2011).

When compositional effects were included in Model 4, the school-level percentage-disadvantaged coefficient was reduced from its Model 3 value but remained significant across all three domains, indicating that school socioeconomic composition is associated with achievement beyond what is explained by school-mean parental education and school-mean ICT resources.

4.3. *Cross level interactions and random slopes*

Model 5 (Cross-Level Interactions, Tables 2–4) tested whether school socioeconomic composition and community type moderated the associations between student characteristics and achievement outcomes. Across all three subject domains, none of the nine cross-level interaction coefficients reached conventional significance ($p < .05$). The strongest hint of moderation was the English X School SES interaction for mathematics, which was marginal ($b = -0.31$, $SE = 0.17$, $p = .074$), suggesting a tendency for English-vs-Spanish gap to narrow in schools with higher concentrations of disadvantaged students; this pattern did not generalize to reading or science. Taken together, the results are consistent with largely additive effects of student and school characteristics on achievement, with school context shifting overall achievement levels rather than reshaping how demographic characteristics relate to achievement. Fig. 4 illustrates the predicted trajectory of achievement by school SES for each immigration group; the nearly parallel lines across the three domains visualize the non-significant interactions.

To more comprehensively examine whether student-level associations with achievement vary across schools, Model 6 (Random Slopes, Tables 2 – 4) allowed the effects of key individual factors (gender, home language, and parental education) to vary across schools. Results for Model 6 appear as the sixth column in Tables 2 – 4 alongside the preceding specifications.

The random-slopes analysis revealed substantial between-school heterogeneity in demographic associations with achievement. Gender slopes varied considerably across schools, indicating that the male disadvantage in reading and the male advantage in mathematics were not uniform: in some schools the gender gap in reading was near zero while in others it exceeded 30 points, with analogous variation in the mathematics gender gap (Fig. 5, top panel).

Home language slopes showed the largest heterogeneity. The English-vs-Spanish achievement gap varied dramatically across schools, from roughly zero in some institutions to exceeding 100 points in others (Fig. 5, bottom panel). This variation likely reflects differences across schools in English-language support programs, bilingual education approaches, and the peer composition of language-minority students.

Within-school parental education slopes showed more modest but still appreciable variation across schools, indicating that while the within-school SES-achievement gradient existed in most schools, its magnitude differed—some schools could better compensate for lower parental education than others.

Fig. 5 visualizes these heterogeneity patterns through violin plots showing the distribution of school-specific associations. Most schools clustered near the pooled mean, but sizable tails are evident in both directions, indicating schools where demographic associations operate quite differently from the national average.

The school intercept variance (τ_{00}) in Model 6 was non-trivial for every domain (reading 1804, mathematics 1006, science 1842), indicating persistent achievement differences between schools even after controlling for measured student and school characteristics and allowing slopes to vary. This residual school variance suggested unmeasured institutional factors (e.g., instructional quality, school leadership, organizational culture) continued to differentiate schools in ways not

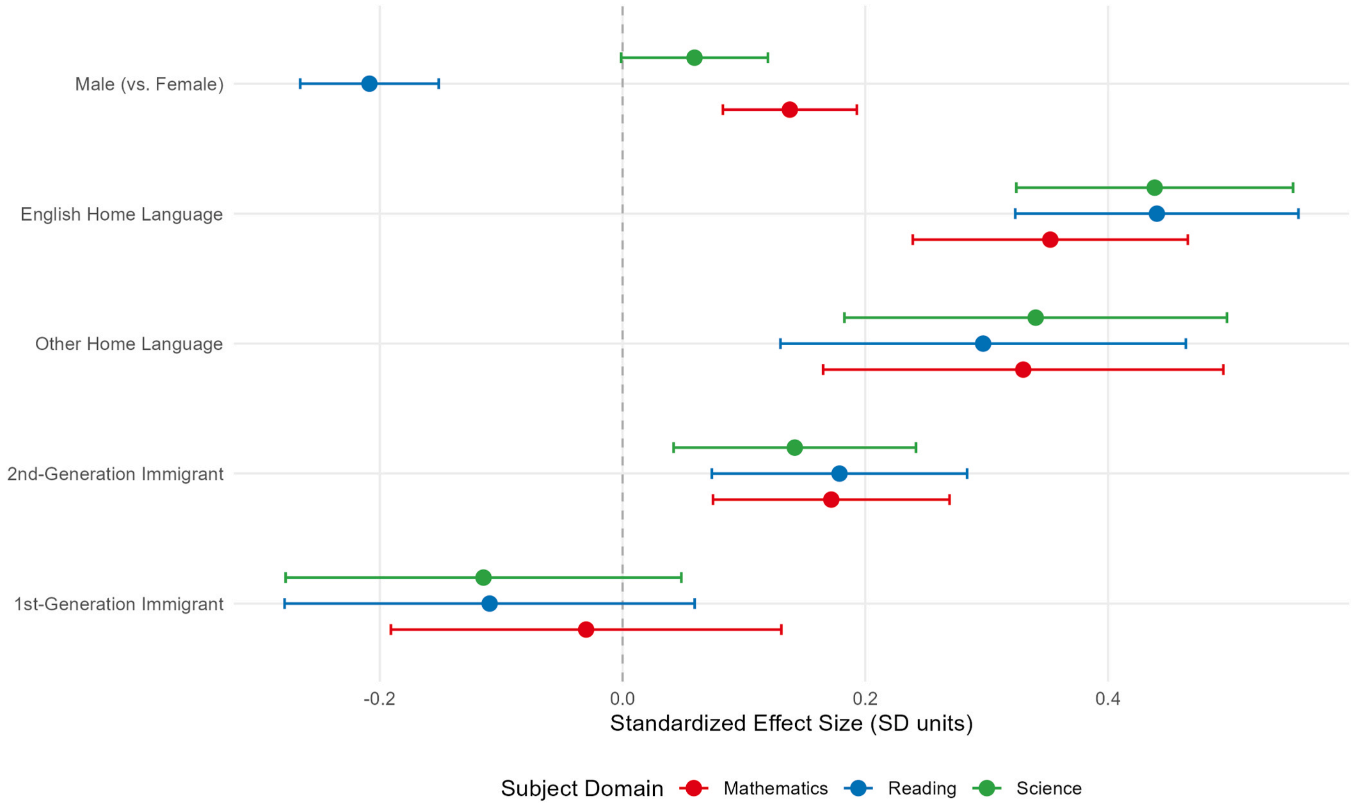

Fig. 2. Student Background Characteristics and Achievement: Standardized Effects Across Domains. Note. Standardized effect sizes were computed by dividing each unstandardized coefficient from the Full Model (Model 3) by the pooled outcome standard deviation for each subject. Error bars show 95% confidence intervals based on Rubin's rules pooling across 10 plausible values X 10 imputations. Reference categories: Gender (Female), Home Language (Spanish), Immigration Status (Native-born).

captured by the variables available in PISA.

Notably, the ICC in Model 6 (0.159 reading, 0.136 mathematics, 0.177 science) declined relative to the Null Model (Model 1) but rebounded modestly from the Model 5 CLI values (0.109 reading, 0.139 mathematics, 0.130 science). This pattern reflects the redistribution of between-school variance from the random intercept to the newly estimated random slopes rather than an overall worsening of fit. Marginal R^2 in Model 6 reached 0.108 (reading), 0.117 (mathematics), and 0.116 (science), modestly lower than the corresponding Model 5 values because some explanatory variance is absorbed by random slopes rather than by fixed effects, a transfer consistent with the intent of Model 6 to surface heterogeneity that a fixed-slope specification would conceal.

5. Discussion

RQ1. : How much variance is at the student vs. school level, and does it differ by subject?

The variance decomposition revealed that 18.9% of reading achievement variance, 24.5% of mathematics variance, and 22.4% of science variance occurred between schools rather than within schools. These findings align with international evidence: the OECD (2023) reports average between-school variance of approximately 29% for mathematics, 26% for science, and 23% for reading across OECD countries in PISA 2022. The U.S. patterns fall at the lower end of these international ranges while preserving similar domain-specific ordering, consistent with Wang et al.'s (2023) systematic review of school-level factors in mathematics achievement.

The domain-specific variance pattern represents a notable finding. Mathematics shows 5.6 %age points more school-level variance than reading and 2.1 points more than science, suggesting somewhat greater institutional sensitivity in mathematics, with science in between. While the practical significance of these differences is open to interpretation, the consistent ordering across model specifications suggests systematic rather than random variation. This pattern contrasts with Korpershoek et al. (2015), who reported broadly similar school-level associations for mathematics and reading. The ordering here is consistent with mathematics' stronger reliance on sequenced coursework, placement, and specialized staffing, while reading development is more diffusely supported across school and out-of-school settings.

Model progression reinforces this interpretation. Moving from null to fully contextual specifications, marginal R^2 rose from near zero to 0.116 in reading, 0.138 in mathematics, and 0.126 in science, while ICCs declined but remained nontrivial (~0.110 reading, ~0.142 mathematics, and ~0.133 science in Model 4). These patterns indicate that institutional characteristics, particularly socioeconomic composition and home resource aggregates, are associated with achievement above and beyond individual background, consistent with Perry and McConney's (2010) compositional-effects framework.

RQ2. : How do student background factors relate to performance across domains?

This study revealed both consistent and domain-specific patterns in how demographic, socioeconomic, and linguistic factors relate to achievement across reading, mathematics, and science.

Gender Patterns. The domain-specific gender pattern (males lower in reading, higher in mathematics, marginal in science) falls within the international PISA 2022 ranges (OECD, 2023) and aligns with Stoet and Geary's (2013) observation that mathematics and reading gender differences vary inversely across countries. Bay and Monseur (2016) reported similar patterns in earlier PISA cycles, suggesting that these domain-specific associations persisted across the pandemic period

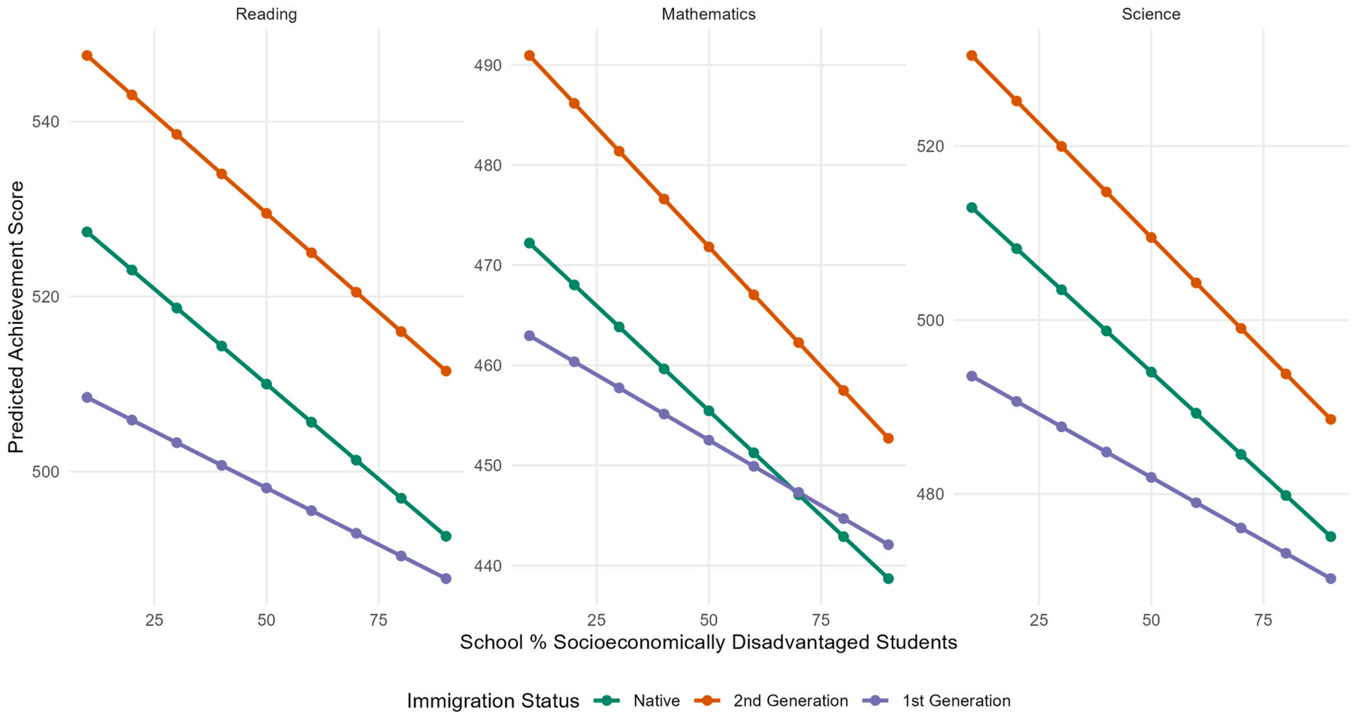

Fig. 3. Compositional Effects: Comparing Individual and School-Level Parental Education Associations. Note. Green bars show within-school associations (individual-level parental education holding school-mean parental education constant; orange bars show between-school associations (school-mean prenatal education holding individual parental education constant). All estimates are from Model 4 (Contextual Effects), pooled via Rubin's rules across 10 plausible values X 10 imputations. Error bars represent 95% confidence intervals. Between-school associations are roughly 3.7–3.8x larger than within-school associations across all three domains. The wider between-school confidence intervals reflect smaller Level−2 sample size (N = 154 schools) and the use of manifest (rather than latent) school means (Lüdtke et al., 2008; Lüdtke et al., 2011).

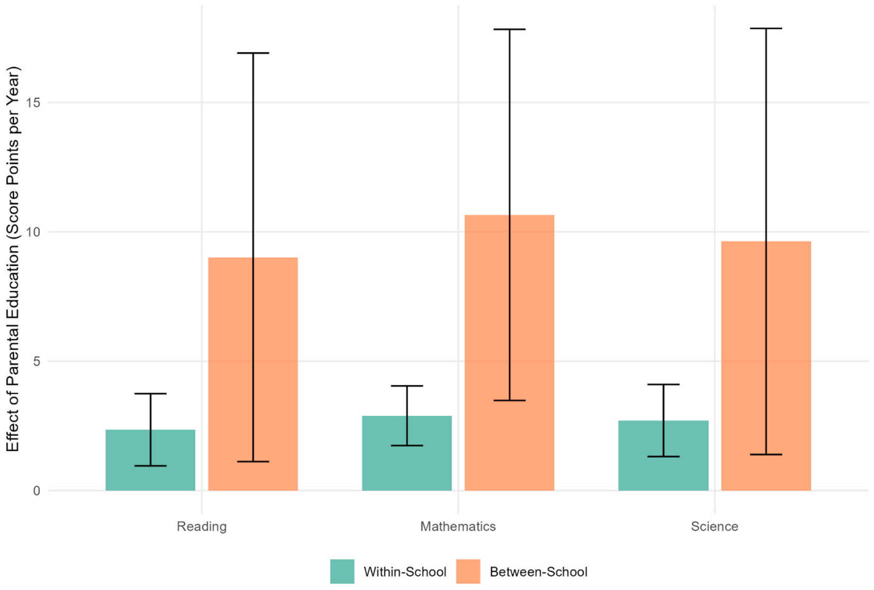

Fig. 4. Predicted Achievement by School Socioeconomic Composition and Immigration Status (Cross-Level Interaction Model). Note. Predicted values derived from Model 5 (Cross-Level Interactions), with other predictors held at reference values: Female for gender, English for home language, the smallest community category for community type, and sample means for parental education, home ICT resources, and school ICT availability. Lines are parallel across immigration status, reflecting the non-significant interactions between immigration status and school socioeconomic composition (all $p > .20$ across the domains).

although the random-slopes results show substantial school-to-school variation around these averages, indicating that institutional factors can amplify or mitigate these gaps rather than producing a single uniform pattern.

Home Language. The English-speaker advantage over Spanish-speaking peers (~43 points in reading and science, ~30 points in mathematics)substantially exceed linguistic-distance effects reported by Borgonovi and Ferrara's (2020) in European contexts, potentially reflecting greater challenges for Spanish-speaking students with U.S. English-language assessments. The pattern of larger language gaps in

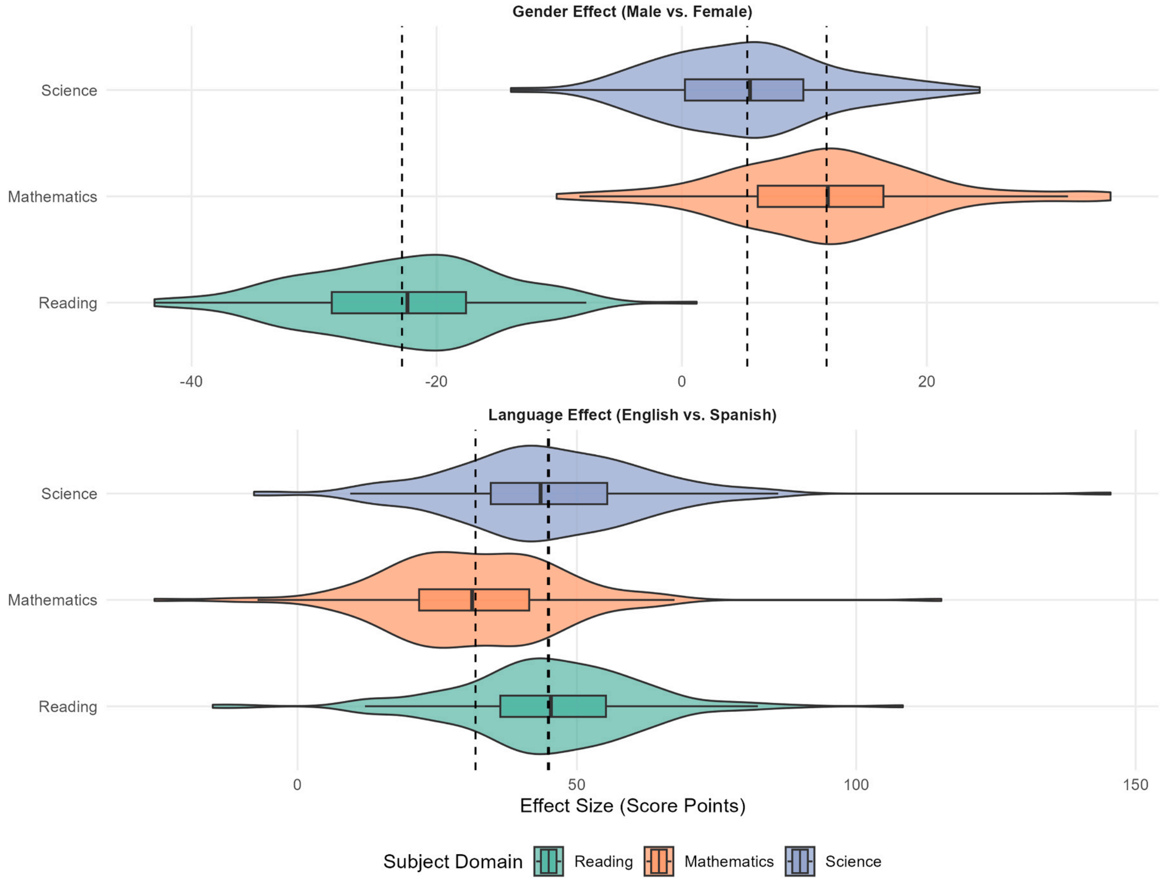

Fig. 5. School-Specific Variation in Gender and Language Effects (Random Slopes). Note. Violin and boxplot distributions represent school-specific effects (the sum of each school's estimated random slope and the corresponding pooled fixed effect) across 154 schools, drawn from a representative fit of Model 6 (Random Slopes). Dashed vertical lines indicate the pooled fixed-effect estimate for each domain. The wide spread of school-specific effects documents substantial heterogeneity in how gender and home language relate to achievement across U.S. schools; across-fit variation is small relative to the displayed school-level heterogeneity.

reading and science than mathematics is consistent with differential linguistic demands across subjects (Curran & Kitchin, 2019), with mathematics' smaller gap suggesting numerical reasoning is less dependent on language-dependent.

Immigration Status. Immigration status revealed intergenerational patterns with appropriately calibrated uncertainty after pooling across plausible values. Second-generation immigrant students consistently outperformed native-born peers across all domains in Model 3: 18.95 points in reading, 15.65 points in mathematics, and 14.74 points in science. These advantages persisted after controlling for home language, parental education, and school contexts simultaneously, and they extend earlier findings by Ghimire (2024) by quantifying second-generation advantages while adjusting for multiple confounders. Substantively, the cross-domain consistency of the second-generation advantage (14 – 19 points) suggests general mechanisms rooted in strong educational values within immigrant-origin families (Kim et al., 2020), combined with native-level English proficiency and familiarity with U.S. schooling conventions.

For first-generation immigrant students, point estimates were in the expected negative direction but confidence intervals spanned zero in all three domains: −11.61 points in reading, −2.73 in mathematics, and −11.91 in science. The wider intervals partly reflect the smaller first-generation subsample (n = 211) and properly-propagated PV uncertainty. These null results should be interpreted as "evidence of uncertainty" rather than "evidence of no effect"; the data are consistent with either a small first-generation disadvantage or no disadvantage once parental education and home language are controlled. Substantively, the qualitative pattern, second-generation advantage clearly above native, first-generation point estimates below native but indistinguishable from it, is consistent with immigrant integration trajectories described in prior work (Crosnoe & Turley, 2011; Kao & Tienda, 1995).

Socioeconomic and Technological Factors. Parental education showed cross-domain-consistent positive associations (approximately 2.5–3.0 points per year), aligning with Wang et al.'s (2023) meta-analytic finding for mathematics and extending it to all three domains. The cross-domain consistency suggests general mechanisms: home learning resources, academic expectations, capacity to navigate educational systems, rather than domain-specific processes. Home ICT resources also showed robust associations (3.2–3.9 points per SD-unit) but smaller than those reported by Ghimire and Mokhtari (2025), possibly because basic ICT access had become more universal by 2022 (reducing its predictive power) or because unmeasured aspects of technology use (digital literacy, purposeful educational engagement, parental mediation) matter more than device access. The persistence of home-ICT effects after controlling for school ICT suggests home technology environments provide benefits not substitutable by school resources.

RQ3. : School Context and Multilevel Effects

School Compositional Effects: ***Systematic Institutional Advantage***. The contextual effects models (Model 4, Tables 2–4) revealed compositional patterns where school-level effects substantially exceeded their within-school counterparts. School socioeconomic composition showed negative associations (−0.41 points per percentage point of disadvantaged students in reading, −0.42 in science, and −0.37 in mathematics), demonstrating how concentrated disadvantage is associated with systematic institutional constraints affecting all students regardless of individual backgrounds, aligning with Liu et al.'s (2020) emphasis on school composition as a critical contextual factor.

School-mean parental education was associated with achievement gains substantially larger than the individual-level (within-school) effects across all domains (Fig. 3). Each additional year of school-mean parental education was associated with gains of 9.02 points in reading, 10.65 points in mathematics, and 9.63 points in science, compared to individual-level effects of 2.35, 2.90, and 2.71 points respectively (Model 4, Tables 2–4), approximately a 3.7–3.8-fold compositional difference, broadly consistent with Perry and McConney's (2010) framework.

Larger still, school-mean home ICT resources were associated with 42.36 points in reading, 42.95 points in mathematics, and 44.81 points in science, approximately eleven-fold the corresponding within-school estimates of 3.60, 3.95, and 3.31 points, respectively (Model 4, Tables, 2–4). These compositional associations extend Chiu's (2020) findings about school technology infrastructure, suggesting that institutional digital capacity is associated with achievement to a greater extent than individual home device access alone, plausibly through coordinated pedagogical integration, technical support, and collaborative learning environments. The wider confidence intervals around these between-school estimates warrant emphasis, however: school-mean coefficients are treated here as manifest aggregates rather than latent school means (Lüdtke et al., 2008; Lüdtke et al., 2011), so the multipliers above should be read as central estimates within a framework that does not fully propagate school-level sampling error.

Domain-specific compositional patterns provided additional insights. Mathematics showed the strongest school-mean parental education association (10.65 points per year), consistent with this domain's highest between-school variance (ICC = 0.245), while science showed the largest school-mean ICT association (44.81 points), perhaps reflecting science's particular reliance on digital simulations, data visualization tools, and interactive resources that schools provide more readily than individual households. Nonetheless, such interpretations warrant cautious reflection. Research grounded in the Big-Fish-Little-Pond framework (e.g., Marsh, 1987) has shown that higher-achieving or resource-rich school environments may simultaneously suppress individual academic self-concept, offsetting some cognitive benefits. Dicke et al. (2018) further demonstrated that certain school-average achievement effects can represent "phantom" compositional influences arising from unmodeled self-concept-achievement dynamics. These counter-perspectives suggest that while institutional advantages may amplify academic outcomes, they could also introduce psychosocial disparities, underscoring the need for integrative models that jointly consider cognitive and affective dimensions of learning.

Cross Level Interaction Results. Despite theoretical justification for cross-level interaction, Model 5 did not detect statistically significant interactions between school socioeconomic composition (or community type) and student-level demographic factors in any of the three subject domains. Points estimates for the nine interaction terms were small and, with a single marginal exception for English-vs-Spanish x School SES in mathematics ($p = .074$), uniformly non-significant. This finding contrasts with Xie and Ma's (2019) identification of significant cross-level interactions in their analysis, potentially reflecting differences in national contexts, measurement approaches, or the specific post-pandemic period examined. Fig. 4 visualizes this null result: predicted achievement declines with rising school disadvantage at essentially the same rate for native, second-generation, and first-generation students, with the three groups offset by intercept differences reflecting main effects. From a policy perspective, the absence of significant moderation supports a "main effects" interpretation of school socioeconomic context: higher-disadvantage schools are associated with lower achievement for students of all demographic backgrounds, rather than disproportionately penalizing any one subgroup.

Random Slopes: Institutional Heterogeneity. While formal cross-level interactions were non-significant, random slopes specification (Model 6) revealed substantial between-school heterogeneity in demographic associations with achievement. Gender slopes varied considerably across schools, home language effects showed even higher variation (school-specific English-vs-Spanish gaps ranging from near zero to over 100 points), and within-school parental-educational slopes varied modestly but significantly (Tables 2 – 4, Model 6; Fig. 5). These heterogeneity patterns represent perhaps the study's most important policy-relevant finding: demographic achievement gaps are not inevitable fixed quantities but vary systematically across schools, consistent with the possibility that school-level practices shape the extent to which demographic background translates into differential achievement. Importantly, this heterogeneity was not captured by the specific Level−2 predictors included in Model 5; the cross-level interactions were uniformly non-significant, yet Model 6 nonetheless documents substantial school-to-school variation. This pattern suggests that the mechanisms producing school-specific demographic gaps are not well represented by aggregate structural variables available in PISA (school SES composition, community size, ICT infrastructure), pointing toward a future research agenda focused on instructional and organizational features not routinely captured in international assessments.

While few studies examine random slopes for demographic factors across subject domains simultaneously, the findings from this study align with Willms's (2010) conclusion that school composition effects vary substantially across institutions. This study extends such work by quantifying heterogeneity for specific demographic factors and demonstrating consistent patterns across three achievement domains.

5.1. *Policy and practice implications*

While acknowledging the correlational nature of the findings, several patterns warrant consideration for educational policy and practice. The substantial institutional heterogeneity revealed through random slopes analysis suggests that some schools narrow demographic achievement gaps while others amplify them, indicating that equity outcomes reflect modifiable institutional practices rather than inevitable demographic differences. Identifying and studying these differentially effective schools could inform system improvement efforts. This finding takes on added weight given that the cross-level interaction tests in Model 5 returned uniformly non-significant coefficients: the school-level structural variables available in PISA (composition, community type, ICT infrastructure) did not, in this sample, account for the school-to-school variation in demographic effects, suggesting that the relevant institutional levers are organizational and instructional rather than purely structural.

The strong home language associations, particularly in reading and science, underscore the importance of robust language support services. The differing patterns between first- and second-generation immigrant students highlight the need for differentiated approaches recognizing both the documented advantages of second-generation students and the more uncertain circumstances of first-generation students, whose point estimates were negative but whose confidence intervals included zero in all three domains once measurement uncertainty was properly propagated. The large compositional patterns observed for parental education and home ICT resources point to the potential significance of addressing concentrated disadvantages and digital inequalities, particularly in mathematics where school factors showed the strongest associations. However, policy decisions should be informed by multiple evidence sources, including experimental studies, as cross-sectional analyses cannot establish causality (Anderson-Cook, 2005).

5.2. *International relevance*

While these findings emerge from U.S. contexts, the analytical framework demonstrates how ecological systems approaches and hierarchical modeling can illuminate educational processes within specific national contexts. The substantial compositional association and institutional heterogeneity observed raise important comparative questions: Do similar patterns emerge in educational systems with different levels of socioeconomic segregation, technology access, and school autonomy?

The specific magnitudes documented—compositional effects approximately 3.7- to 3.8-fold for parental education and eleven-fold for ICT resources—provide benchmark for international comparison, with the caveat that these multipliers are central estimates for manifest-aggregate framework and would likely be reported with tighter confidence bounds under a latent-aggregate specification (Lüdtke et al., 2008; Lüdtke et al., 2011). Researchers examining other educational systems can assess whether similar patterns occur in contexts with different levels of socioeconomic segregation, school autonomy, or technology infrastructure. The institutional heterogeneity revealed through random slopes analysis raises a further comparative question: Do some national systems show more uniform demographic relationships across schools, and if so, what organizational or instructional features rather than structural composition might explain this consistency?

Systems with greater residential integration or more equitable resource distribution might show reduced compositional patterns, while systems with heightened segregation might demonstrate larger ones. Applying parallel analytical frameworks across diverse contexts could identify system-level features that either amplify or mitigate compositional advantages, informing evidence-based strategies for promoting educational equity globally.

5.3. *Limitations and future directions*

Several important limitations shape the interpretation of these findings. First, the cross-sectional design prevents causal inferences about the relationships observed (Anderson-Cook, 2005; Maxwell & Cole, 2007). While the associations identified are statistically robust and theoretically meaningful, establishing causality requires experimental or longitudinal designs (Gustafsson, 2013). This limitation particularly affects interpretations of pandemic-related patterns, as pre-post comparison would be necessary to isolate COVID−19 effects.

Second, hierarchical linear modeling assumptions (e.g., normally distributed random effects, linear relationships) may oversimplify complex patterns (Goldstein, 2011). The two-level structure omits classroom-level factors (teacher quality, instructional practices) that mediate between individual and institutional influences (Lüdtke et al., 2009).

Third, omitted variable bias may affect the estimates, as many potentially important variables (e.g., teaching practices, classroom climate, curriculum implementation, peer networks, and school leadership) were unavailable. The large compositional associations observed, may partly reflect unmeasured selection processes or neighborhood factors (Dicke et al., 2018).

Fourth, the United States did not meet PISA's minimum school response rate threshold in the 2022 assessment cycle (51% before replacement and 63% after, vs. the 65% and 85% targets; OECD, 2023). The student response rate (80%) met the target, and the OECD nonresponse bias analyses indicated that measured characteristics were broadly balanced, but the possibility of bias cannot be excluded. Schools that choose not to participate may differ systematically from participating in ways that affect estimates of school-level variance and compositional effects.

Fifth, school-level aggregates (school-mean parental education and school-mean ICT resources) were treated as manifest aggregates rather than latent school means. Lüdtke et al., 2008; Lüdtke et al., 2011 document that latent-aggregate approaches correct for sampling error in school-level estimates and typically produce tighter confidence intervals, although point estimates are usually comparable. This study retained the manifest approach because it is tractable in the {lme4} framework used for PV-as-imputation pooling and because the Level−2 N (154 schools) is at the lower end of the range where latent aggregation offers substantial improvement. Readers should nonetheless interpret our compositional-effect magnitudes as central estimates within a framework that does not explicitly propagate school-level sampling error.

Sixth, the School ICT Availability (ICTSCH) index exhibits a pronounced ceiling effect in the U.S. PISA 2022 sample: 83% of schools cluster at the maximum observed value, with only 17% of schools showing lower values. ICTSCH coefficients should therefore be interpreted as capturing a contrast between the modal well-resourced schools and a smaller subset of lower-scoring schools, rather than as a smooth dose-response relationship across the full range of school ICT infrastructure. This limitation affects all three models incorporating ICTSCH (Models 3–6).

Seventh, the broad categorizations used here (e.g., English vs. Spanish speakers, native vs. immigrant) may mask important within-group variations, and U.S.-focused analysis limits generalizability to contexts with different educational structures.

Future research should explore the mechanisms underlying the compositional effects observed in this study. Experimental and quasi-experimental designs could test whether interventions targeting school composition or resources causally affect achievement outcomes. Additionally, longitudinal studies tracking students overtime could help establish temporal precedence and differentiate between selection and socialization effects in school compositional relationships.

The substantial heterogeneity in school-specific relationships suggests that value of mixed-methods research combining statistical analyses with in-depth case studies. Identifying the specific policies, practices, and cultural features that characterize schools with smaller demographic gaps could generate hypotheses for subsequent experimental testing. Finally, comparative international research could explore whether the patterns observed in the U.S. context appear in different educational systems, potentially identifying policy approaches associated with more equitable outcomes across diverse national contexts.

6. Conclusion

This study demonstrates how ecological systems theory and multilevel modeling illuminate complex relationships between individual characteristics and institutional contexts in shaping achievement. Three insights emerge.

First, between-school variance was substantial and domain-specific, with mathematics highest (24.5%) followed by science 22.4% and reading 18.9%, a pattern that challenges the assumptions that school effects operate uniformly across subject-specific institutional approaches to addressing achievement gaps.

Second, random slopes analysis reveals substantial institutional heterogeneity in demographic achievement relationships across schools. This heterogeneity demonstrates that educational disparities are not uniform but vary with institutional contexts, suggesting potential for school-level practices to shape equity outcomes.

Third, compositional effects findings demonstrate that school-level aggregates (school-mean parental education, school-mean ICT resources) show substantially stronger associations with achievement than their individual-level counterparts, roughly 3.7x for parental education and 11x for home ICT resources, although with wider confidence intervals reflecting the smaller Level−2 sample and the use of manifest rather than latent aggregates. These patterns illustrate how advantages may compound when concentrated within schools, with the appropriate interpretive caution noted in the Limitations.

Together, these findings extend beyond Bronfenbrenner's ecological framework and compositional-effects by quantifying the relative magnitude of individual and school-level influences and by documenting substantial institutional heterogeneity in how demographic factors relate to achievement, pattens invisible to singl-moel specifications.

While cross-sectional design precludes causal conclusions, the multilevel relationships described here provide a benchmark for future comparative work examining whether similar patterns appear in educational systems with different levels of socioeconomic segregation, technology access, or school autonomy.

CRediT authorship contribution statement

Nirmal Ghimire: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or non-for-profit sectors.

Declaration of Competing Interest

The author has no known conflict of interest to disclose.

Acknowledgements

The author completed all aspects of this study independently.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.stueduc.2026.101620.

Data Availability

The PISA data collected from 2022 cycle were used to conduct this study. They can be accessed through the OECD international database, https://www.oecd.org/pisa/data/2022database/

References

Anderson-Cook, C. M. (2005). Experimental and Quasi-experimental designs for generalized causal inference. *Journal of the American Statistical Association, 100*(470), 708. https://doi.org/10.1198/jasa.2005.s22

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3–25. https://doi.org/10.1037/amp0000191

Atasever, U., Huang, F. L., & Rutkowski, L. (2025). Reassessing weights in large-scale assessments and multilevel models. *Large-Scale Assessment in Education, 13*(9). https://doi.org/10.1186/s40536-025-00245-y

Avvisati, D. (2020). The measure of socio-economic status in PISA: A review and some suggested improvements. *Large-Scale Assessments in Education, 8*(8), 1–37. https://doi.org/10.1186/s40536-020-00086-x

Bates, D., Machler, 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

Bay, A., & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-Scale Assessment in Education, 4*(1), 1–16. https://doi.org/10.1186/s40536-015-0015-x

Borgonovi, F., & Ferrara, A. (2020). Academic achievement and sense of belonging among no-native-speaking immigrant students: The role of linguistic distance. *Learning and Individual Differences, 81*, Article 101911. https://doi.org/10.1016/j.lindif.2020.101911

Borgonovi, F., & Pokropek, M. (2021). The evolution of the association between ICT use and reading achievement in 28 countries. *Computers and Education Open, 2*, Article 100047. https://doi.org/10.1016/j.caeo.2021.100047

Bronfenbrenner, U., & Morris, P. A. (2007). The biological model of human development. In R. M. Lerner, & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 793–828). Jonh Wiley & Sons, Inc.

Camacho, D. E., & Fuligni, A. J. (2015). Extracurricular participation among adolescents from immigrant families. *Journal of Youth and Adolescence, 44*, 1251–1262. https://doi.org/10.1007/s10964-014-0105-z

Chiu, M. S. (2020). Exploring models for increasing the effects of school information and communication technology use on learning outcomes through outside-school use and socioeconomic status mediation: The ecological techno-process. *Educational Technology Research and Development, 68*, 413–436. https://doi.org/10.1007/s11423-019-09707-x

Chiu, M., & Chow, B. W.-Y. (2015). Classmate characteristics and student achievement in 33 countries: Classmates' past achievement, family socioeconomic status, educational resources, and attitude toward reading. *Journal of Educational Psychology, 107*(1), 152–169. https://doi.org/10.1037/a0036897

Crosnoe, R., & Turley, R. N. L. (2011). K-12 educational outcomes of immigrant youth. *The Future of Children, 21*(1), 129–152. https://doi.org/10.1353/foc.2011.0008

Curran, F. C., & Kitchin, J. (2019). Why are the early elementary race/ethnicity test score gaps in science larger than those in reading or mathematics? National evidence on the importance of language and immigration context in explaining the gap-in-gaps. *Science Education, 103*(3), 477–502. https://doi.org/10.1002/sce.21491

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology, 110*(8), 1112–1126. https://doi.org/10.1037/edu0000259

Donnelly, R., & Patrinos, H. A. (2021). Learning loss during Covid-19: An early systematic review. *Prospects, 51*, 601–609. https://doi.org/10.1007/s11125-021-09582-6

Drescher, J., Podolsky, A., Reardon, S. F., & Torrance, G. (2022). The geography of rural educational opportunity. *The Russell Sage Foundation Journal of the Social Sciences, 8*(3), 123–149. https://doi.org/10.7758/RSF.2022.8.3.05

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

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

Gajderowicz, T., Jakubowski, M., Kennedy, A., Kjeldsen, C.C., Patrinos, H.A., & Strietholt, R. (2025). *The learning crisis: Three years after COVID-19*. *arXiv*. https://doi.org/10.48550/arXiv.2501.01260.

Ghimire, N. (2024). Understanding disparities: Examining demographic, socioeconomic, and linguistic impacts on U. S. students' outcomes in reading, math, and science. *Advance*. https://doi.org/10.31124/advance.24226158.v1

Ghimire, N., & Mokhtari, K. (2025). Mapping connections across layers: Interrelationships between key factors and 15-year-old's reading scores. *Reading Psychology, 43*(2), 195–227. https://doi.org/10.1080/02702711.2024.2432882

Ghimire, N., & Regmi, S. (2020). Do school and school district influence students' performance on civics EOC assessment? A multilevel analysis. *Florida Journal of Educational Research, 58*(7), 3–20. https://doi.org/10.62798/QJZN4830

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester, UK: John Wiley & Sons. https://doi.org/10.1002/9780470973394

Grund, S., Lüdtke, O., & Robitzsch, A. (2021). *mitml:Tools for multiple imputation in multilevel modeling. R package version 0.4-5 [R package]*. https://doi.org/10.32614/CRAN.package.mitml.

Gubbels, J., Swart, N. M., & Groen, M. A. (2020). Everything in moderation: ICT and reading performance of Dutch 15-year-olds. *Large-Scale Assessment in Education, 8*(1). https://doi.org/10.1186/s40536-020-0079-0

Gustafsson, J. E. (2013). Causal inference in educational effectiveness research: A comparison of three methods to investigate effects of homework on student achievement. *School Effectiveness and School Improvement, 24*(3), 275–295. https://doi.org/10.1080/09243453.2013.806334

Kao, G., & Tienda, M. (1995). Asian-Americans as model minorities? A look at their academic performance. *American Journal of Education, 103*(2), 121–159. https://doi.org/10.1086/444094

Kim, Y., Mok, S., & Seidel, T. (2020). Parental influences on immigrant students' achievement-related motivation and achievement: A meta-analysis. *Educational Research Review, 30*, Article 100327. https://doi.org/10.1016/j.edurev.2020.100327

Korpershoek, H., Kuyper, H., & van der Werf, G. (2015). The relation between students' math and reading ability and their mathematics, physics, and chemistry examination grades in secondary education. *International Journal of Science and Mathematics Education, 13*(5), 1013–1037. https://doi.org/10.1007/s10763-014-9534-0

Kuhfeld, M., Soland, J., Lewis, K., & Morton, E. (2022). *The pandemic has had devastating impacts on learning. What will it take to help students catch up?* United States of America: Brookings Institution. 〈https://coilink.org/20.500.12592/kv8kz9〉.

Kyriakides, L., Creemers, B. P., Panayiotou, A., & Charalambous, E. (2020). *Quality and equity in education: Revisiting theory and research on educational effectiveness and improvement*. Routledge.

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Economic Policy Institute.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Liu, H., Van Damme, J., Gielen, S., & Van Den Noortgate, W. (2020). School processes mediate school compositional effects: Model specification and estimation. *British Educational Research Journal, 41*(3), 423–447. https://doi.org/10.1002/berj.3147

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthen, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13*(3), 203–229.

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2×2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods, 16*(4), 444–467. https://doi.org/10.1037/a0024376

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*(2), 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001

Lumley, T. (2010). *Complex surveys: A guide to analysis using R. John.* Wiley and Sons.

Maldonado, J. E., & De Witte, K. (2022). The effect of school closure on standardized student test outcomes. *British Educational Research Journal, 48*(1), 49–94. https://doi.org/10.1002/berj.3754

Mang, J., Kuchenhoff, H., Meinck, S., & Prenzel, M. (2021). Sampling weights in multilevel modeling: An investigation using PISA sampling structures. *Large-Scale Assessment in Education, 9*(6). https://doi.org/10.1186/s40536-021-00099-0

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*(3), 280–295. https://doi.org/10.1037/0022-0663.79.3.280

Martinez-Abad, F., Gamazo, A., & Rodriguez-Conde, M.-J. (2020). Educational data mining: Identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation, 66*, Article 100875. https://doi.org/10.1016/j.stueduc.2020.100875

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods, 12*(1), 23–44. https://doi.org/10.1037/1082-989X.12.1.23

Meinck, S., Fraillon, J., & Strietholt, R. (2022). *The impact of the COVID-19 pandemic on education: International evidence from the responses to educational disruption survey (REDS)*. UNESCO.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*(2), 177–196. https://doi.org/10.1007/BF02294457

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

NCES. (2019). *Program for International Student Assessment (PISA) 2018: Weighting*. National Center for Education Statistics. ⟨https://nces.ed.gov/surveys/pisa/2018technotes-4.asp⟩.

OECD. (2018). *The Resilience of Students with an Immigrant Background: Factors that Shape Well-being*. OECD Publishing. https://doi.org/10.1787/9789264292093-en

OECD. (2023). *PISA 2022 results: Ready to learn, thrive and contribute, I*. OECD Publishing.

Organization for Economic Cooperation and DevelopmentOECD). (2023). *PISA 2022 Database [Data set]*. ⟨https://oecd.org/en/data/datasets/pisa-2022-database.html⟩.

Perry, L. B., & McConney, A. (2010). Does the SES of school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record, 112*(4), 1137–1162. https://doi.org/10.1177/016146811011200401

R Core Team. (2026). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. ISBN 3-900051-1-07-0. ⟨http://www.R-project.org/⟩.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, 1*. SAGE.

Rubin, D. B. (1987). The Calculation of Posterior Distributions by Data Augmentation: Comment: A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations When Fractions of Missing Information Are Modest: The SIR Algorithm. *Journal of the American Statistical Association, 82*(398), 543–546. https://doi.org/10.2307/2289460

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: issues in secondary analysis and reporting. *Educational Researcher, 39*(2), 142–151. https://doi.org/10.3102/0013189X10363170

Schleicher, A. (2023). *PISA 2022: Insights and interpretations*. OECD Publishing.

Sparks, S.D. (2023). *U.S. teenagers decline in global test of math, but hold steady in reading, science. Education Week*. ⟨https://www.edweek.org/leadership/u-s-teenagers-decline-in-global-test-of-math-but-hold-steady-in-reading-science/2023/12⟩.

Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 Years of PISA Data. *PLoS ONE, 8*(3), Article e57988. https://doi.org/10.1371/journal.pone.0057988

Sulak, T. (2016). School Climate and academic achievement in suburban schools. *Education and Urban Society, 48*, 672–684. https://doi.org/10.1177/0013124514514541465

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

van Dijk, J. (2005). *The deepening divide: Inequality in the information society*. SAGE.

von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). What are the plausible values and why are they useful?. In *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, 2* pp. 9–36) IEA/ETS Research Institute.

Wang, X. S., Perry, L. B., Malpique, A., & Ide, T. (2023). Factors predicting mathematics achievement in PISA: A systematic review. *Large-Scale Assessments in Education, 11*, 24. https://doi.org/10.1186/s40536-023-00174-8

Wentzel, K. R., & Ramani, G. B. (2016). *Handbook of social influences in school contexts*. Taylor & Francis.

Willms, J. D. (2010). School composition and contextual effects on student outcomes. *Teachers College Record, 112*(4), 1008–1037. https://doi.org/10.1177/016146811011200408

Xie, C., & Ma, Y. (2019). The mediating role of cultural capital in the relationship between socioeconomic status and student achievement in 14 economies. *British Educational Research Journal, 45*(4), 838–855. https://doi.org/10.1002/berj.3528