JOURNAL OF CULTURAL ANALYSIS AND SOCIAL CHANGE

Journal of Cultural Analysis and Social Change
2025, 11(1), 500-512
ISSN: 2589-1316

Hierarchical Linear Modeling of Grade 8 Math Achievement in TIMSS: Explaining Learning Gaps in Egypt

Sarah El-Gerby[1*], Sherine Al-Shawarby[2]

[1] *PhD Candidates, Economics and Political Science, Cairo University.*
Lecturer Assistant, Economics and Public Finance Department, Kafr-Elsheikh University, Egypt, Address: Department of Economics and Public Finance, Kafr-Elsheikh University, Kafr-Elsheikh University Avenue, Egypt. Email: sarahelgerby@aucegypt.edu, ORCID : 0000-0002-3511-1910
[2] *Professor of Economics, Department of economics, Nile University (seconded from Cairo University Economics department), Egypt, sherine.alshawarby@feps.edu.eg*

Corresponding Author: sarahelgerby@aucegypt.edu

Citation: El-Gerby, S. & Al-Shawarby, S. (2026). Hierarchical Linear Modeling of Grade 8 Math Achievement in TIMSS: Explaining Learning Gaps in Egypt, *Journal of Cultural Analysis and Social Change*, 11(1), 500-512. https://doi.org/10.64753/jcasc.v11i1.3893

Published: January 01, 2026

ABSTRACT

The global education agenda has evolved from the Millennium Development Goals (MDGs), which focused primarily on universal access, to the Sustainable Development Goals (SDG 4), which place paramount emphasis on quality education. This study examines the determining factors of mathematics achievement in Egypt to identify barriers to SDG 4. While Egypt has achieved significant success in expanding educational access, international assessments reveal a persistent crisis in learning outcomes. This study investigates the determinants of math achievement among Egyptian Grade 8 students using data from the Trends in International Mathematics and Science Study (TIMSS) 2019. Applying a two-level hierarchical linear model (HLM) to a sample of 7,210 students, we disentangle the effects of student-level socioeconomic status (SES) and school-level resources. Our findings confirm the dominant role of socioeconomic disparities, revealing a nonlinear relationship where the benefits of wealth and parental education diminish at higher levels. In contrast, conventional school-level inputs show limited explanatory power. The results lead to the conclusion that achieving Sustainable Development Goal (SDG4) in Egypt requires redistributive policies that address structural socioeconomic barriers alongside targeted school-based reforms.

Keywords: education quality; TIMSS; Egypt; socioeconomic status (SES); hierarchical linear model (HLM); mathematics achievement; SDG 4

INTRODUCTION

The measurement of education quality has evolved significantly, moving away from input-based proxies such as enrollment rates and school attendance, which often fail to reflect actual learning outcomes (Hanushek, 1986). In response, a growing literature emphasizes student performance on International Large-Scale Assessments (ILSAs) like the Trends in International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS), and the Programme for International Student Assessment (PISA) as more reliable and comparable metrics of human capital quality across countries(Dixon, 2019; Hanushek & Wößmann, 2007). This shift is grounded in robust evidence demonstrating that cognitive skills, rather than mere years of schooling, are the primary drivers of individual earnings and national economic growth (Hanushek & Wößmann, 2012; Schoellman, 2011).

This paradigm is reflected in the global development agenda. While the Millennium Development Goals (MDGs) successfully expanded access to education, the Sustainable Development Goals (SDGs), particularly SDG4, now prioritize inclusive, equitable, and quality education (UNDP, 2018). This is critical in contexts like Egypt, which has achieved near-universal enrollment, with gross enrollment rates rising from 91% in 2014/15 to 104% in 2019(Economics., (n.d.); Schwab, 2018; Schwab & Sala-i-Martin, 2015). However, this quantitative success masks a profound quality crisis. Egypt has consistently ranked near the bottom globally on education quality metrics, placing 133rd out of 137 countries in the 2017/18 Global Competitiveness Report(Schwab, 2018). While in the TIMSS 2019 cycle, only 26% and 23% of Grade 8 students met intermediate benchmarks in math and science, respectively. Alarmingly, Nearly 99% of students succeeded in passing Egypt's Grade 9 national exams proficiency standards, suggesting a concerning disconnect between national assessment standards, actual learning outcomes and students' progress through levels (Education, 2023; Mullis et al., 2020). A central debate in the economics of education concerns the relative influence of socioeconomic status (SES) versus school-level resources on student achievement. Seminal works, from the Coleman Report in the United States to the Heyneman-Loxley effect in lower-income countries, have framed this discussion. Contemporary research suggests that as countries develop, the influence of household SES often surpasses that of school inputs in predicting achievement (Baker et al., 2002; Coleman, 1966; Heyneman & Loxley, 1983). However, the application of this research to Egypt remains limited. While studies on Egyptian education often focus on gender equity, access, dropout ratios, and macroeconomic returns (AlBagoury, 2024; Education, 2023; Elbadawy, 2015; Salehi-Isfahani et al., 2014), nationally representative analyses of the hierarchical determinants of learning outcomes are scarce. A comprehensive review of the literature reveals that large-scale, quantitative analyses leveraging internationally comparable data like TIMSS to investigate the determinants of math achievement are particularly limited. (Badr, 2012; Badr et al., 2012a, 2012b). A handful of exceptions exist but they are few, leaving a critical knowledge gap

This study aims to address this gap by providing one of the first comprehensive analyses of the determinants of math achievement in Egypt using nationally representative TIMSS 2019 data. We employ a two-level Hierarchical Linear Model (HLM) with data from 7,210 Grade 8 students to disentangle the effects of household SES, teacher characteristics, and school-level resources. Our research tests two central hypotheses derived from the literature:

H1: Student-level socioeconomic status (SES) factors (parental education, household wealth) will have a stronger influence on math achievement than school-level factors (teacher qualifications, resources).

H2: The relationship between SES and achievement is nonlinear, exhibiting diminishing returns at higher levels of wealth and parental education.

The remainder of this paper is structured as follows: Section 2 reviews the theoretical and empirical literature. Section 3 describes the data and methodology. Section 4 presents the results, and Section 5 discusses the findings and their policy implications. Section 6 concludes by noting the study's limitations and suggesting avenues for future research.

LITERATURE REVIEW

Defining and Measuring Education Quality

The definition of education quality is contested, primarily between two perspectives. The economic approach, grounded in Human Capital Theory, evaluates quality through measurable outcomes such as test scores, returns to schooling, and labor market performance (Aghion et al., 2009; Becker, 2002; Schoellman, 2011). In contrast, a humanist approach which is a concept in the development of human resource prospective and emphasizes the educational process itself, including student engagement, holistic development, and equity (Stonkuvienė et al., 2021). In international policy, the economic perspective has gained dominance, with ILSAs serving as key proxies for education quality.

Human Capital Theory posits that investments in education enhance individual productivity, leading to higher earnings and economic growth (Becker, 2002; Carneiro et al., 2010). A growing body of evidence confirms that cognitive skills, measured by standardized tests, are a stronger predictor of economic outcomes than schooling attainment alone (Hanushek & Wößmann, 2012; Schoellman, 2011).). International Large-Scale Assessments (ILSAs), such as the Trends in International Mathematics and Science Study (TIMSS), provide critical, comparable data to monitor educational quality and diagnose the systemic inequalities that hinder progress toward SDG4. For example, (Schoellman 2011) finds a strong correlation between returns to schooling and performance on standardized tests, further emphasizing that quality, not just quantity, is the true driver of human capital.

As Egypt only participated in the Grade 8 TIMSS assessments until 2019, [Table 1] summarizes Egypt's math scores since 2003, highlighting persistent gaps despite slight improvements

Table 1 EGYPT MATH SCORE IN THE 4 ROUNDS COMPARED TO THE INTERNATIONAL AVERAGE SCORE

Year	Average Score (SD)	International Average
2003	406 (3.5)	467
2007	391 (3.6)	500 *
2015	392 (4.1)	500 *
2019	413 (5.2)	500 *

* Source: IEA's Trends in International Mathematics and science study TIMSS 2019 downloaded from HTTP// timss2019.org/download

TIMSS Scale Center-point1

As illustrated in [Table 1], Egypt's performance in TIMSS has remained consistently below the international average, highlighting persistent structural challenges despite slight improvements. This pattern underscores that the focus must shift from access to the quality and effectiveness of education.

The **Baker et al. (2002)** replication of the HL effect, using HLM, found that in most countries—regardless of income level—SES variables now outperform school-level variables in predicting achievement. They argue that economic development has diminished the HL effect by elevating household influence relative to institutional factors. At the same time, (Boca et al., 2018; Carneiro et al., 2010)argue that early-life investments yield the highest returns in human capital formation, reinforcing the importance of targeted education interventions in early childhood—particularly for disadvantaged households.

Socioeconomic Status versus School-Level Determinants

Early landmark studies, particularly the Coleman Report (1966), concluded that family background—measured by parental education, household income, and home learning conditions—was more predictive of academic outcomes than school resources in The U.S. This view has long informed education policy in high-income countries, often summarized by the phrase "money doesn't matter", as a pointer to minimal effect of government spending on school (Coleman, 1966). While Heyneman and Loxley (Heyneman & Loxley, 1983) argued that school resources were more influential in low-income countries. Later, Hanushek and Luque (2003) found that contrary to the conventional view, school resources are not disproportionately important for student achievement in low-income contexts (Hanushek & Luque, 2003), reinforcing the complexity of the debate over whether public education funding or family background matters more. A seminal replication by Baker et al. 2002 using HLM found that SES now outperforms school-level variables in most countries, regardless of income level, suggesting economic development elevates household influence.

Becker's Household Production Theory provides a micro-foundation, positing that families convert time and resources into child outcomes (Becker, 2002). Empirical evidence consistently shows that children from lower-SES backgrounds face persistent disadvantages in cognitive development (Cooper & Stewart, 2021; Ermisch & Francesconi, 2001). This pattern holds in the Middle East and North Africa (MENA) region, where SES is a pivotal driver of educational inequality (Coşkun & Karadağ, 2023; Wiberg et al., 2024). In Egypt, while girls often outperform boys academically, this advantage does not translate into labor market parity, underscoring the role of non-academic, sociocultural mediators (Badr et al., 2012b).

The Role of School-Level and Systemic Factors

Despite the dominant role of SES, school-level factors remain significant, particularly in contexts of systemic disadvantage. Evidence from both high- and low-income countries consistently identifies teacher quality as one of the most influential school-level determinants of student achievement. Studies using TIMSS 2015 data indicate that students taught by subject-qualified teachers achieved significantly higher test scores—an effect especially pronounced among female students taught by female teachers. A study using TIMSS data to assess the impact of teachers' college majors on students' achievements they found that Teachers enhance student achievement in natural science subfields aligned with their college majors. Approximately half of this effect is attributable to teaching practices, with the majority explained by teachers' specialized preparation in science education. Professional development for teachers also contributes to improved student performance, with evidence indicating that teacher training correlates strongly with improved outcomes across diverse contexts , researchers found that certified teachers positively influenced student achievement, suggesting policy investments should focus on dual-capacity building in both mathematical knowledge and instructional methods(Darling-Hammond, 2000; Outlook, 2012; Wallace, 2025) Similarly, World Bank (2018) cross-country studies highlight that teacher subject knowledge and teaching practices—not merely credentials or years of service—are key predictors of learning outcomes. Other

1 The TIMSS achievement scale was established in 1995 based on the combined achievement distribution of all countries that participated in TIMSS 1995. To provide a point of reference for country comparisons, the scale Center Point of 500 was located at the mean of the combined achievement distribution.

research using PISA data has provided empirical evidence that school-level economic, social, and cultural status (ESCS) significantly predicts the probability of academic success for disadvantaged students(Agasisti et al., 2021).

However, the effectiveness of school inputs is often mediated by systemic conditions. In Egypt, challenges such as large class sizes—which rose from 41 to 49 students on average between 2010 and 2020—fragmented teacher training, and a highly centralized administration hinder the impact of resources (Program(UNDP), 2021). Overcrowding reduces student-teacher interaction, limits formative assessment, and lowers classroom engagement (Konstantopoulos, 2008). Moreover, Egypt's education system remains highly centralized, with minimal school-level autonomy. (Bruns et al., 2013) show how weak accountability systems and lack of school-level discretion diminish the return on education investments.

Evidence suggests the relationship between spending and outcomes is nonlinear; increased funding improves learning only up to a certain threshold, beyond which efficiency becomes paramount (Vegas & Coffin, 2015; Vegas & Umansky, 2005). Studies in Egypt have found that per-student public spending has only a minimal effect on achievement, highlighting these systemic inefficiencies (Husseiny & Amin, 2018).

Contribution of This Study

While prior research on Egyptian education has examined trends in expenditure or analyzed enrollment indicators in isolation, few studies have systematically linked these factors to learning outcomes using advanced statistical modeling. This paper fills this critical gap by employing a two-level HLM to investigate the hierarchical determinants of math achievement. Our analysis provides one of the first comprehensive assessments of the contributors to basic education quality in Egypt, offering evidence-based insights for optimizing resource allocation and designing effective policies to address the nation's learning crisis.

Data and Analytical Strategy

This study examines the mathematics achievement determinants among Grade 8 students in Egypt using TIMSS 2019 data. Given the *nested nature* of the data—students nested within classrooms and schools—the study uses two levels Hierarchical Linear Model (HLM), which accounts for clustering and enables robust estimation of student- and school-level influences. The analysis also incorporates plausible values, stratified sampling weights, and jackknife standard errors to ensure valid statistical inference. The section begins by outlining the modeling strategy, followed by a description of the dataset, variables, and estimation approach.

Analytical Framework and Modeling Strategy

To examine how student-level socioeconomic characteristics and teacher- or school-level factors influence educational outcomes, the study employed a two-level HLM. This framework accounts for TIMSS's multistage cluster sampling—where students are nested within classes and schools—by modeling variation at both the individual students (Level 1) and between-school or classroom-level factors (Level 2). HLM mitigates the risk of underestimated standard errors and inflated Type I errors, thereby ensuring valid inference across levels. The model was estimated twice using two different analytical tools in R: the widely used *lme4* package (Bates et al., 2014) and the more recent *WeMix* package (Bailey, 2023), which is optimized for large-scale assessments like TIMSS and allows direct integration of plausible values, a key feature of (ILSA) such as TIMSS.

Given the nested structure of the data, HLM was employed. Hence, it accounts for data clustering, which is critical to avoid underestimated standard errors and inflated Type I errors. This methodology allows for a concurrent analysis of relationships at the student and school levels, ensuring robust estimates.

A foundational model's suitability, we estimated a baseline (null) empty model to compute the Intra-class Correlation Coefficient (ICC) - Equation (1), quantifying the share of variance in achievement attributable to differences between schools. The hierarchical linear model's fit and assumptions were tested to ensure its validity for applying HLM where:

$$ICC = \sigma^2 \times \frac{1}{\sigma^2 + \tau_{00}^2} \qquad \text{(Equation1)}$$

Where:

σ^2 : The variance between groups (e.g., schools), which is the variance of the random effect.

τ_{00}^2 : The residual variance (within-group variance) is the variance of the residuals or the error term. The denominator here represents the total variance (both between and within schools). A value of the ICC between 0.05 and 0.2 is sufficient to apply the HLM model.

We found that the adjusted Intra-class Correlation ICC= 0.3718. According to The ICC was estimated at 0.372, indicating that 37.2% of the variance in math achievement is attributable to differences between schools, according to(Boulifa & Kaaouachi, 2022; Coşkun & Karadağ, 2023; Muthén & Satorra, 1989). Practically, this suggests that school significantly contributes to explaining the students' performance in mathematics as well.

To address the complex structure of TIMSS data, the study incorporated all the 5 plausible values and applied appropriate weighting, analysis model for our present study: .

$$M_{mij} = \beta_{0ij} + \beta_m S_{mij} + \gamma_m SCH_{mj} + e_{ij} + U_{mj} \quad \text{(Equation 2)}$$

Where:

M_{mij}: Matrix of the five plausible values of math scores (indexed by *m*) for student *i* in school *j*

S_{mij}: Vector of student-level socioeconomic covariates (e.g., gender, parental education) of student *i* in school *j* - (Level 1)

SCH_{mj}: Vector of school- and teacher-level covariates (e.g., class size, teacher qualifications, school location) for school *j* (Level 2)

e_{ij}: Level 1 residual error, assumed to be independently and identically distributed across students

U_{mj}: Level 2 random effect, representing unobserved school-level influences on outcome *m*, assumed to be independent of e_{ij} and across schools (Grilli, Pennoni et al., 2016)

Equation 2 specifies our full model, where the outcome variable is the set of five plausible values for each student's math score. Student-level covariates (S_mij) and teachers' school-level covariates (SCH_mj) are included to assess cross-level interactions, while the residual terms (e_ij and U_mj) capture unobserved student- and school-level effects, respectively.

jackknife replication methods (JRR) were applied to correct standard errors, following TIMSS conventions. Jackknife's **variance and standard error** were computed using Equations 3 and 4.

Standard errors were corrected using jackknife replication (75 replicates) in the lme4 method, with variance and standard errors computed following TIMSS conventions. Variance and standard error were computed using Equations 3 and 4 in lme4.

$$Var_{Jack}(\widehat{\theta}) = \frac{K-1}{K} \sum_{K=1}^{K} (\hat{\theta}^K - \bar{\theta})^2 \quad \text{Equation 3}$$

$$SE_{jack}(\hat{\theta}) = \sqrt{Var_{jack}}(\hat{\theta}) \quad \text{Equation 4}$$

Where:

- K: Number of jackknife replicates (number of clusters).
- $\hat{\theta}^K$: Estimate from k[th] replicate (dropping one cluster).
- $\widehat{\theta}$: Full-sample estimate.

Data and Variables

The dataset comes from Egypt's TIMSS 2019, covering 7,210 Grade 8 students across 169 schools. TIMSS uses a stratified, multistage cluster sampling methodology to ensure national representativeness. The dependent variable is mathematics achievement, represented by five plausible values derived through Item Response Theory (IRT)(Mullis et al., 2016), capturing student competencies across number, probability, algebra, geometry, and applied mathematics (Cognitive Domains).

- **Level 1(Student Level) Variables**

Key variables include gender (BSBG01), a wealth index constructed by the author based on household assets, and home possessions (NWealthI) see [table 2] below.

Table 2 WEALTH INDEX BY PERCENTILE GROUP

		Frequency	Percent	Valid Percent	Cumulative Percent
	1st quintile	7506	17.4	17.4	17.4
	2nd	11688	27.0	27.0	44.4
	3ed	7056	16.3	16.3	60.7
	4th	9408	21.7	21.7	82.4
	5th	7602	17.6	17.6	100.0
Valid	Total	43260	100.0	100.0	

Note: *The distribution across wealth percentiles is relatively balanced, with no single group dominating the sample. Percentiles 2 and 4 together comprise 48.7% of students, indicating a broad middle segment. This reflects Egypt's socioeconomic structure, especially in governmental schools. Notably, students from wealthier households tend to underperform in math. contrary to international patterns, suggesting that higher household wealth does not necessarily translate into better academic outcomes in the Egyptian context.*

(Badr, 2012) found that superior academic performance observed in single-sex language schools, relative to their Arabic-language counterparts, and this is strongly linked to the socioeconomic characteristics of their students family . In early 2000s Egypt, household spending on shadow education (non-school) exceeded government education expenditure on schools, with persistently high private tutoring costs (Rahmouni, 2025)

Parental Education (BSDGEDUP), reflecting the highest level attained by either parent, see [table 3]

Table 3 parental education

Parents' Highest Education Levels					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	University or Higher	13014	30.1	30.1	30.1
	Post-secondary but not University	9546	22.1	22.1	52.1
	Upper Secondary	3936	9.1	9.1	61.2
	Some Primary, Lower Secondary or No School	8826	20.4	20.4	81.7
	Don't Know	7038	16.3	16.3	97.9
	999	900	2.1	2.1	100.0
	Total	43260	100.0	100.0	

Note: *Only 30.1% of students have parents with university or higher education, while 9.1% report parents who completed upper secondary school. A significant portion of students (20.4%) come from families with only primary, lower secondary, or no formal education, and 16.3% responded "Don't know," highlighting gaps in students' awareness of their parents' educational attainment.*

Teacher/School-Level Variables (Level 2),

Variables include teacher gender (BTBG02), class size (BTBG10), whether the teacher majored in mathematics (BTBG05A), perceptions of malnutrition's impact on learning (BTBG13B), parental support (BTBG06He), instructional time per day (BCBG06B), school location (BCBG5B), perceived teacher shortages (BCBG13BA), and the school's socioeconomic composition (BCDGSBC). These were drawn from student, teacher, and principal questionnaires. Definitions and coding are in following table. Descriptive statistics are presented in table

Table 4 DEFINITIONS OF THE VARIABLES WITH THEIR CODES

Dependent variable (5 plausible values)	
BSMMAT01 BSMMAT02 BSMMAT03 BSMMAT04 BSMMAT05	-1st Plausible value in mathematics -2nd Plausible value in mathematics -3rd Plausible value in mathematics -4th Plausible value in mathematics -5th Plausible value in mathematics
Independent variables	
First level variables (Students)	
-BSBG01 -NWealthI - BSDGEDUP	-Student Gender (with female as a reference group) - Wealth index - Parents' Highest Education level
Second level (Teacher and school level data)	
-BTBG02 -BTBG10 -BTBG05A -BTBG13B - BTBG06He	-GEN\SEX of teacher -Class size -Teacher's major is Math - How much lack of nutrition limit teaching - parental support for their kids
BCBG05B BCBG06B BCBG13BA BCDGSBC	- Immediate area of school location -Total instructional time/Mins -How much is your school affected by shortage or inadequacy of teachers specialized in math- School Composition by Socioeconomic Background Parental education - School Composition by Socioeconomic Background

Data Processing and Estimation Tools

Prior to estimation, the dataset underwent several preprocessing steps as follows:

Missing Data Treatment: Multiple imputation via the MICE package in R was used to address missing data under the Missing at Random (MAR) assumption (Bouhlila & Sellaouti, 2013; Soley-Bori, 2013).

Table 5 PERCENTAGE OF MISSING DATA FOR USED VARIABLES

Variable	**Percentage missing**
Student gender	0.2%

Private tutoring	36.1%
Parents' highest education level	2%
Class size	5%
Teacher major	0.14%
How lack of nutrition limits teaching	2.83%
Parental Support	0.4%
Total instructional time	10.3%
School composition by SES	16.6%
-Calculated by the Authors	

Variable Selection and Recoding: Variables aligned with the research framework were selected and recoded as necessary. Descriptive statistics (means, standard deviations, minimum and maximum values) were computed for all variables. Descriptive statistics are presented in table 6. The descriptive statistics for the main variables, including mathematics scores, socioeconomic indicators, and school-level characteristics highlight notable patterns and anomalies in the TIMSS 2019 Egypt dataset.

Table 6 descriptive statistics of the variables

Variable	N	Minimum	Maximum	Mean	Std. Dev.
1st Plausible Value Mathematics	43260	60.015	752.332	412.765	92.201
2nd Plausible Value Mathematics	43260	63.152	711.141	411.944	94.241
3rd Plausible Value Mathematics	43260	78.197	720.868	412.181	94.477
4th Plausible Value Mathematics	43260	89.05	745.819	410.689	94.8
5th Plausible Value Mathematics	43260	89.764	732.219	411.932	94.094
Student Gender	43249	0	1	0.47	0.499
Wealth Index	43260	1	5	2.95	1.372
Parental Education	43260	1	999	23.43	142.21
Teacher Gender	167	0	1	0.63	0.485
Class Size	162	18	83	44.78	12.099
Teacher Math Major	159	0	1	0.2	0.402
Malnutrition Limits Teaching	165	1	3	2.05	0.607
Parental Support	165	1	3	1.897	0.712
School Location	169	1	999	37.43	185.037
Instructional Time	151	180	540	335.85	56.422
Math Teacher Shortage	169	1	99	2.43	7.526
School SES Composition	138	1	3	2.54	0.716

Note: *Key observations include:*

From the table we can noticed that the five mathematics plausible values show consistent means around 411–413 with substantial variability (SD ≈ 92–95), indicating wide performance dispersion among students. The Student-level variables such as gender and wealth index display balanced distributions, whereas parental education exhibits extreme variability, suggesting possible confusing answers by the students about their parents. Teacher- and school-level variables show limited variation and moderate dispersion due to Egypt's public school system is somehow highly standardized and homogeneous, particularly evident in class size and instructional time.

Diagnostic Testing: Multicollinearity (via VIF), heteroscedasticity (Breusch-Pagan), and autocorrelation (Durbin-Watson) were tested. No major violations were detected.

The HLM models were estimated using both the lme4 and WeMix packages in R While lme4 is widely used and supports jackknife replication (Bates et al., 2014), WeMix is optimized for large-scale assessments such as TIMSS and directly incorporates plausible values and school-level weights (Bailey, 2023). The use of both tools enhances robustness and informs methodological choice for future research. All five plausible values were included to address measurement error in the IRT scaling process , while employing jackknife repeated replication (JRR) to correct standard errors, ensuring valid population-level inference.

RESULTS AND INTERPRETATION

This section presents and interprets the findings from the two-level hierarchical linear model estimating the determinants of mathematics achievement among Grade 8 students in Egypt. The analysis proceeds in four stages. First, Section 4.1 estimates the variance components from the empty model to justify the use of multilevel techniques. Second, Section 4.2 summarizes key coefficients and methodological differences between estimation strategies. Third, Sections 4.3 and 4.4 interpret results in terms of substantive student- and school-level predictors and compare outcomes across models.

Between-School Variation: The Empty Model

As shown in Equation 1, the ICC, derived from the empty model (M0), was estimated at 0.372. This indicates that 37.2% of the total variance in students' mathematics achievement is because of differences between schools. This level of between-school variation substantiates the use of hierarchical modeling approach. These results are consistent with (Boulifa & Kaaouachi, 2022; Muthén & Satorra, 1989), and support the relevance of the Heyneman-Loxley hypothesis (Heyneman & Loxley, 1983), which highlights the importance of school-level factors in shaping educational outcomes in developing contexts. The subsequent analysis examines how student and school-level variables contribute to this variance.

Methodological Comparison of Multilevel Model Estimates

We compare the estimates generated by the two modeling approaches—WeMix and lme4 with jackknife standard errors (lme4+JK)—to assess consistency in coefficient magnitude and statistical precision, and interpretability.

Table 7 presents coefficient estimates and standard errors for across both models. The **WeMix** model (using mixPV function), optimized for large-scale assessment data, and incorporates plausible values and sampling weights directly, resulting in smaller standard errors and improved sensitivity to statistically significant relationships. Conversely, the **lme4+JK** model often yields larger coefficients, reflecting a more flexible estimation process that may highlight stronger substantive associations despite more conservative standard errors.

Table 7 Results table

Term	lmer + JK Estimate (SE)	lmer + JK p-value	mixPV Estimate (SE)	mixPV p-value
(Intercept)	2.162305 (2.345)	0.359392	0.532290383 (0.270)	0.0491 *
BSBG01girl	0.805456 (0.276)	0.004722 **	0.162135773 (0.044)	0.0002 ***
NWealthI.L	-0.54552 (0.168)	0.001725 **	-0.109561332 (0.031)	0.0006 ***
NWealthI.Q	-1.1785 (0.203)	1.55E-07 ***	-0.232808399 (0.0274)	0.0000 ***
NWealthI.C	-0.33116 (0.166)	0.049376 *	-0.064254078 (0.0259)	0.0182 *
NWealthI^4	-0.24395 (0.143)	0.091938.	-0.047845317 (0.0259)	0.0684.
BSDGEDUP.L	0.3884 (0.166)	0.021812 *	0.078933335 (0.0308)	0.0124 *
BSDGEDUP.Q	-0.39891 (0.158)	0.013445 *	-0.079338589 (0.0299)	0.0107 *
BSDGEDUP.C	-0.65018 (0.190)	0.001023 **	-0.128139000 (0.02995)	0.0002 ***
BSDGEDUP^4	-1.50696 (0.234)	1.02E-08 ***	-0.296575589 (0.0238)	0.0000 ***
BTBG02Male	-0.10013 (0.401)	0.803465	-0.014266464 (0.1467)	0.9225
BTBG10	-0.02985 (0.083)	0.720191	-0.006090712 (0.0060)	0.3167
BTBG05Ayes	0.368007 (2.271)	0.871909	0.099411338 (0.0986)	0.3219
BTBG06He2	-1.57609 (7.225)	0.827907	-0.236680660 (0.1169)	0.0487 *
BTBG06He3	-1.6003 (5.401)	0.767831	-0.405291919 (0.1897)	0.0328 *
BTBG13BNot at all	0.2931 (4.034)	0.94227	-0.049263325 (0.2079)	0.8127
BTBG13BSome	-0.20713 (5.357)	0.969259	-0.156821541 (0.1483)	0.2905
BCBG06B	-0.00183 (0.112217)	0.987021	-0.004221602 (0.0048)	0.3762
BCBG13BANot at all	-0.92921	0.350425	-0.134469911	0.1594

	(0.989)			
BCBG13BA some or a little	-0.99425 (12.579)	0.937218	-0.145251180	0.5577
BCDGSBC More Disadvantaged	-0.44835 (0.626)	0.476087	-0.085052517 (0.0654)	0.1952
BCDGSBC Neither More Affluent nor More Disadvantaged	-0.25872 (0.632)	0.68351	-0.0390205 (0.0935)	0.6770

*Significance levels: *** p<0.001, ** p<0.01, * p<0.05, p<0.1*
Standard errors are in parentheses.
Source: Author's calculations based on TIMSS 2019 Grade 8 dataset (Egypt).

Summary of key significant predictors with comparison between LME4 and WEMIX models is in Appendix table A1.

The **consistency in sign and statistical significance** across both methods strengthens confidence in the results. However, the WeMix model's smaller standard errors make it more sensitive in detecting significant effects. The lme4+JK model, though more conservative in its standard error estimates, tends to yield larger coefficients, which may imply stronger substantive effects.

Taken together, these findings suggest that while both models are methodologically valid, their outputs should be interpreted with an understanding of their respective strengths. *WeMix offers precision; lme4+JK provide flexibility and interpretive clarity*. This dual-method approach offers a more comprehensive understanding of both the statistical reliability and practical significance of the factors driving performance disparities, enhancing the robustness of the overall analysis.

Interpretation First Level Predictors

This section interprets the estimated effects of key predictors using both the lme4+JK and WeMix results, highlighting the direction, magnitude, and robustness of these effects. The key findings are as follows:

- **Gender Differences in Achievement:** Girls consistently outperformed boys in mathematics achievement. The WeMix model estimates a gender advantage of 0.16 points, while the lme4 model reports a higher effect of 0.81 points. Although modest in absolute terms, this consistent gap aligns with regional patterns reported in Badr et al. (2012) and Schwab (2018), and suggests that academic gains for girls, despite limited translation into labor market gains.
- **Parental education** also emerged as a significant predictor of student performance. Parental education shows a modest positive linear effect but a negative quadratic component, implying diminishing returns at higher parental education levels**.** In both models, each additional level of parental education is associated with improved achievement—an increase of 0.08 points in the WeMix model and 0.39 points in the lme4 model. However, the presence of a significant negative quadratic term in both models (WeMix: –0.08; lme4: –0.40) implies diminishing returns at higher levels of parental education. These findings are broadly consistent with the human capital framework (Becker, 2002; Ermisch & Francesconi, 2001), yet it also suggests that beyond a certain threshold, active parental involvement and intervention may be more critical to student success than parental education.
- **Household wealth exhibited a counterintuitive negative association with achievement.** The linear effects were negative in both models (WeMix: –0.11; lme4: –0.55), and the quadratic terms also followed this pattern (WeMix: –0.23; lme4: –1.18), suggesting a nonlinear relationship where the benefits of wealth taper off—or even decline—at higher levels. This result contrasts with global expectations and may reflect context-specific factors such as disparities in school quality or weak links between household affluence and educational engagement. Notably, the influence of wealth on performance was smaller than that of parental education; the *magnitude of the wealth effect is relatively small,* especially in the WeMix model.

Interpretation of School-Level Predictors

Among school-level predictors, parental support --as teachers disclose-- appears to have a modest but statistically significant influence on student performance, as estimated by the multilevel models. Higher reported levels of parental support at the school level are associated with gains of 0.24 to 0.41 points across models. These contextual effects underscore the importance of family–school engagement in the learning process, consistent with findings from (Kareshki & Hajinezhad, 2014) in similar MENA region contexts.

Comparison Across Models

Both the WeMix and lme4 models produce broadly consistent results in terms of direction and statistical significance. However, the WeMix estimates tend to have smaller standard errors, reflecting its native handling of plausible values and sampling weights. In contrast, the lme4 model often yields larger coefficient magnitudes, particularly for student-level predictors such as gender, wealth, and parental education. School-level effects, while generally smaller in both models, appear more conservative in lme4. These differences suggest that software choice can influence both the perceived precision and practical relevance of findings, with potential implications for policy interpretation.

Validation and Robustness Checks

To ensure the reliability of the model estimates, several diagnostic tests and robustness checks were conducted. Variance Inflation Factor (VIF) values for all predictors were well below the threshold of 10, *indicating no serious multicollinearity*. The Durbin-Watson test produced a value of 1.88, suggesting no significant autocorrelation in residuals. Homoscedasticity was confirmed via the Breusch-Pagan test (p = 0.7274), failing to reject the null hypothesis of constant error variance.

Methodological comparisons further reinforced robustness. The **mixPV** model yielded smaller standard errors due to its native support for plausible values and sample weights, increasing the sensitivity for detecting significant effects. In contrast, the **lme4+JK** approach, though more conservative with larger standard errors, offered transparent interpretability and cross-validation.

These diagnostics and cross-model comparisons enhance confidence in the and reinforce the appropriateness of the HLM approach to capture the nested structure of TIMSS data. Nonetheless, despite these robustness checks, certain methodological and data limitations remain, as discussed next.

Policy Implications in a Broad Context

Our results have direct and profound policy implications. The finding that SES is a primary constraint suggests that policies like Egypt's conditional cash transfer program, (*Takaful and Karama)*, are well-targeted. However, the nonlinear effects imply that these programs should be designed to provide strong support at the lower end of the SES spectrum. Furthermore, school-based reforms must be "poverty-sensitive," focusing on strategies that are effective in disadvantaged contexts, such as improving teaching quality in core subjects and providing targeted learning support, rather than blanket investments in infrastructure.

The stark contrast between Egypt's increased education spending and stagnant TIMSS scores highlights a critical issue of spending efficiency over volume. Our evidence suggests that without complementary policies to mitigate socioeconomic disadvantages, investments in schools alone are unlikely to yield the desired learning gains.

Despite increased education spending—from 4.8% to 5.6% of GDP—and a constitutional commitment to 6%, Egypt continues to face stagnant learning outcomes. This suggests that policy attention must shift from spending volume to spending effectiveness. Key policy priorities should include:

- Modernizing curricula to focus on core competencies and STEM integration
- Expanding school-based health and nutrition interventions for disadvantaged communities
- Scaling up family-oriented programs, such as conditional cash transfers (e.g., *Takaful and Karama*), parental literacy initiatives, and community engagement programs that empower families to support learning at home

More broadly, this study highlights the value of International Large-Scale Assessments (ILSA) like TIMSS not merely as benchmarking tools, but as critical diagnostic instruments for guiding evidence-based policy.

To promote equitable, high-quality education for all, Egypt must adopt a ***dual-track strategy***: one that strengthens vulnerable households while investing in systemic improvements to school infrastructure, teaching, and learning conditions. Only through integrated, data-driven policymaking can achievement gaps be narrowed and the potential of recent education reforms fully realized.

To accelerate progress toward Sustainable Development Goal 4 (Quality Education), we recommend the creation of *a global knowledge-sharing platform* to facilitate peer learning and policy transfer. This platform should support the exchange of evidence-based practices in:

1. *Curriculum Reform:* Competency-based models and STEM integration.
2. *Teacher Professional Development*: Effective training and retention strategies.
3. *Learning Assessment*: Equitable, innovative measurement tools.

A possible implementation mechanism would include:

- A *digital repository*, hosted by UNESCO or the Global Education Monitoring (GEM) Report, featuring case studies, toolkits, and adaptable policy templates;
- *Regional peer learning networks*, such as MENA-focused working groups, to contextualize global insights;

We urge the UN to allocate funding for this platform under its *SDG Acceleration Toolkit*, prioritizing participation from low- and middle-income countries.

LIMITATIONS AND FUTURE RESEARCH

Despite the strengths of this study provides robust and policy-relevant insights, some limitations should be acknowledged. The reliance on self-reported data may introduce measurement error, and while multiple imputation was used, the possibility of data not being Missing at Random (MAR) could bias results. Furthermore, the cross-sectional nature of TIMSS data prevents causal inference. Future research should employ longitudinal designs to track student progress over time and explore the causal mechanisms behind the negative wealth-achievement relationship observed in Egypt. Qualitative studies could also illuminate why higher parental education and wealth show diminishing returns, investigating home learning environments and parental engagement strategies.

Summary Interpretations

Overall, the findings underscore that student-level socioeconomic factors significantly influence learning outcomes in Egypt. This dual impact affirms the value of a multilevel analytical perspective and aligns with prior research suggesting that, while family background remains a strong determinant of achievement, improving school quality is essential to reducing educational inequality (Hanushek & Luque, 2003; Bouhlila & Sellaouti, 2013).

A consistent, though modest, *gender effect* was also observed, with girls outperforming boys in mathematics. This aligns with regional patterns noted by Badr et al. (2012), suggesting that cultural or motivational factors may be at play, even if this academic advantage does not yet translate into labor market parity

Despite the significant between-school variance, none of the specific school-level predictors (teacher major, instructional time, class size) reached statistical significance in our models.

This does not imply these factors are unimportant; rather, it suggests a relative homogeneity in the quality of inputs across the Egyptian public school system, or that their effects are mediated by the strong socioeconomic composition of schools. This finding challenges a pure Heyneman-Loxley interpretation and reinforces the conclusion that simply increasing standard school resources may be insufficient without addressing underlying socioeconomic inequalities. This aligns with earlier work by Hanushek and Luque (2003), which found that resource increases in developing countries do not guarantee better outcomes.

Appendix

A 1 SUMMARY OF KEY SIGNIFICANT PREDICTORS – COMPARISON BETWEEN LME4+JK AND WEMIX MODELS

Variable	Lme4 Estimate + JKse	Interpretation (lmer + JK)	mixPV Estimate	Interpretation (mixPV)
BSBG01girl (Gender)	0.805 **(0.276)**	Girls score 0.805 higher	0.16213 (0.044)	Being a girl score 0.162 higher
NWealthI.L (Wealth, Linear)	-0.546 **(0.168)**	Wealth ↑, score ↓ by 0.546	-0.110 (0.031)	Wealth ↑, score ↓ by 0.110
NWealthI.Q (Wealth, Quadratic)	-1.179 **(0.203)**	Wealth ↑, score ↓ by 1.179	-0.233 (0.0274)	Wealth ↑, score ↓ by 0.233
BSDGEDUP.L (Parental Education)	0.388 **(0.166)**	Parental edu ↑, score ↑ by 0.388	0.079 (0.0308)	Parental edu ↑, score ↑ by 0.079
BSDGEDUP.Q (Parental Education)	-0.399 **(0.158)**	Parental edu ↑, score ↓ by 0.399	-0.079 **(0.0299)**	Parental edu ↑, score ↓ by 0.079
BTBG06He2 (Parental Support: Medium)	-1.576 **(7.225)**	Med. parental support ↓ score by 1.576	-0.237 **(0.1169)**	Med. parental support ↓ score by 0.237
BTBG06He3 (Parental Support: Low)	-1.600 **(5.401)**	Low parental support ↓ score by 1.600	-0.405 **(0.1897)**	Low parental support ↓ score by 0.405

Data Availability Statement

The datasets analyzed during the current study are available from the corresponding author upon reasonable request. The sources of the data used in all tables are calculated by the authors from data published on the World Bank website: https://timss2019.org/international-database/ accessed on 1 September 2022

Acknowledgment: We extend our heartfelt gratitude to Professor Tarek Moursi of Cairo University, whose expertise, vision, and guidance were invaluable during this research. His inspiration helped shape the topic and methodology, and his contributions continue to resonate in our work. We dedicate this research to his memory..”

REFERENCES

Agasisti, T., Avvisati, F., Borgonovi, F., & Longobardi, S. (2021). *What school factors are associated with the success of socio-economically disadvantaged students? An empirical investigation using PISA data. Social Indicators Research*, *157*(2), 749-781.

Aghion, P., Boustan, L., Hoxby, C., & Vandenbussche, J. (2009). *The causal impact of education on economic growth: evidence from US. Brookings papers on economic activity*, *1*(1), 1-73.

AlBagoury, S. H. (2024). Female Education and Economic Growth in Egypt: An Empirical Study. In *Technological Innovations for Business, Education and Sustainability* (pp. 297-308). Emerald Publishing Limited.

Badr, M. (2012). *School effects on educational attainment in Egypt*. https://www.econstor.eu/handle/10419/65450

Badr, M., Morrissey, O., & Appleton, S. (2012a). *Determinants of educational attainment in MENA*.

Badr, M., Morrissey, O., & Appleton, S. (2012b). *Gender differentials in maths test scores in MENA countries*.

Baker, D. P., Goesling, B., & LeTendre, G. K. (2002). *Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman-Loxley effect" on mathematics and science achievemen*t. *Comparative Education Review*, *46*(3), 291-312.

Becker, G. S. (2002). Human capital. *The concise encyclopedia of economics*, *2*.

Boca, D. D., Monfardini, C., & See, S. G. (2018). *Government Education Expenditures, Pre-Primary Education and School Performance: A Cross-Country Analysis*.

Bouhlila, D. S., & Sellaouti, F. (2013). Multiple imputation using chained equations for missing data in TIMSS: a case study. *Large-scale Assessments in Education*, *1*, 1-33. https://link.springer.com/article/10.1186/2196-0739-1-4

Boulifa, K., & Kaaouachi, A. (2022). The relationship between the school resources index; gender; age and mathematics achievement in TIMSS 2019 survey: Multilevel analysis. *Procedia Computer Science*, *201*, 738-745.

Bruns, B., O'Gara, C., Filmer, D., Hanushek, E., Puryear, J., Sandefur, J., & Schmelkes, S. (2013). Schooling Is Not Education! Using Assessment to Change the Politics of Non-learning. *CGD Report*.

Carneiro, P., Dearden, L., & Vignoles, A. (2010). *The economics of vocational education and training*. *International Encyclopedia of Education. Amsterdam: Elsevier*.

Coleman, J. S. (1966). *Equality of educational opportunity [summary report* (Vol. 1). US Department of Health, Education, and Welfare, Office of Education.

Cooper, K., & Stewart, K. (2021). Does household income affect children's outcomes? A systematic review of the evidence. *Child Indicators Research*, *14*(3), 981-1005.

Coşkun, B., & Karadağ, E. (2023). *The effect of student and school characteristics on TIMSS 2015 science and mathematics achievement: The case of Türkiye*. *Journal of Pedagogical Research*, *7*(1), 203-227.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education policy analysis archives*, *8*, 1-1.

Dixon, A. (2019). *The education crisis: Being in school is not the same as learning*. In: The World Bank. https://www. worldbank. org.

Economics., T. ((n.d.)). *Egypt - Gross enrolment ratio, primary and lower secondary, both sexes (%).* . https://tradingeconomics.com/egypt/gross-enrolment-ratio-primary-and-lower-secondary-both-sexes-percent-wb-data.html

Education, G. P. f. (2023). *Pre-University Education sector analysis. Egypt. 2021*. Retrieved Jan26 from https://www.globalpartnership.org/content/education-sector-analysis-egypt-2021

Elbadawy, A. (2015). Education in Egypt: Improvements in attainment, problems with quality and inequality. *The Egyptian labor market in an era of revolution*, *127*.

Ermisch, J., & Francesconi, M. (2001). Family matters: Impacts of family background on educational attainments. *Economica*, *68*(270), 137-156.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of economic literature*, *24*(3), 1141-1177.

Hanushek, E. A., & Luque, J. A. (2003). Efficiency and equity in schools around the world. *Economics of education Review*, *22*(5), 481-502.

Hanushek, E. A., & Wößmann, L. (2007). The role of education quality for economic growth. *The World Bank*.

Hanushek, E. A., & Wößmann, L. (2012*). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation*. *Journal of economic growth*, *17*(4), 267-321.

Heyneman, S. P., & Loxley, W. A. (1983). The distribution of primary school quality within high-and low-income countries. *Comparative Education Review*, *27*(1), 108-118.

Husseiny, I. A. E., & Amin, K. Z. (2018). Pre-university education outputs in Egypt: does money matter? *International Journal of Education Economics and Development*, *9*(3), 210-235.

Kareshki, H., & Hajinezhad, Z. (2014). A Multilevel Analysis of the Role of School Quality and Family Background on Students' Mathematics Achievement in the Middle East. *Universal Journal of Educational Research*, *2*(9), 593-602.

Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from Project STAR. *The Elementary School Journal*, *108*(4), 275-291.

Mullis, I. V., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. In: Boston College, TIMSS & PIRLS International Study Center.

Mullis, I. V., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2016). *TIMSS 2015 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. ERIC.

Muthén, B., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In *Multilevel analysis of educational data* (pp. 87-99). Elsevier.

Outlook, A. E. (2012). *Promoting Youth Employment. Pocket Edition, African Development Bank (AfDB)*. In: OECD Development Centre, UN Development Programme (UNDP), OECD Publishing, Paris.

Program(UNDP), U. N. D. (2021). *Egypt Human Development Report 2021* (0305-9049). https://www.undp.org/egypt/egypt-human-development-report-2021

Rahmouni, M. (2025). *Book Review: Shadow Education in Africa: Private Supplementary Tutoring and Its Policy Implications Bray Mark*. In: SAGE Publications Sage UK: London, England.

Salehi-Isfahani, D., Hassine, N. B., & Assaad, R. (2014). Equality of opportunity in educational achievement in the Middle East and North Africa. *The Journal of Economic Inequality*, *12*(4), 489-515.

Schoellman, T. (2011). Education quality and development accounting. *The Review of Economic Studies*, *79*(1), 388-417.

Schwab, K. (2018). The global competitiveness report 2018.

Schwab, K., & Sala-i-Martin, X. (2015). World Economic Forum's Global Competitiveness Report, 2014-2015. Retrived from: http://reports. weforum. org/global-competitiveness-report-2014-2015/. Switzerland: World Economic Forum,

Soley-Bori, M. (2013). Dealing with missing data: Key assumptions and methods for applied analysis. *Boston University*, *4*(1), 1-19.

Stonkuvienė, I., Juškevičienė, A., & Šabanovas, S. (2021). The Concept of quality of education: a view from Lithuania. ICERI2021 Proceedings,

UNDP. (2018). *Goal 4 Quality Education*. http://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-4-quality-education/targets/

Vegas, E., & Coffin, C. (2015). When education expenditure matters: An empirical analysis of recent international data. *Comparative Education Review*, *59*(2), 289-304.

Vegas, E., & Umansky, I. (2005). Improving teaching and learning through effective incentives. *Incentives to improve teaching*, 1.

Wallace, A. K. (2025). *Mathematics Teacher Certification Status and Student Achievement: A Quantitative Analysis*. [PhD, Xavier University of Louisiana]. XULA digital commons.

Wiberg, M., Laukaityte, I., & Rolfsman, E. (2024). The association between attitudes towards mathematics, students' background and TIMSS mathematics achievement. *European Journal of Mathematics and Science Education*, *5*(1), 13-26.

