

Reframing the Reasoning Gap: Cognitive-Domain Achievement Profiles and Their Predictors Among Eighth-Grade Students in Türkiye Across TIMSS 2019 and 2023

Başak Barak^a, Betül Barut^b

Received : 13 July 2026
Revised : 2 September 2026
Accepted : 30 September 2026
DOI : 10.26822/iejee.2026.467

^aCorresponding Author: Başak Barak, Anadolu University, Department of Science and Mathematics Education
E-mail: bbarak@anadolu.edu.tr
ORCID: <https://orcid.org/0000-0001-5536-7625>

^b Betül Barut, Anadolu University, Department of Science and Mathematics Education
E-mail: byayan@anadolu.edu.tr
ORCID: <https://orcid.org/0000-0002-6196-9654>

Abstract

Eighth-grade mathematics achievement in Türkiye is often framed around a presumed weakness in higher-order reasoning, yet the relative standing of the TIMSS cognitive domains has rarely been tested inferentially or across time. Using the nationally representative TIMSS 2019 ($N = 4,077$) and 2023 ($N = 4,925$) Grade 8 samples for Türkiye, we profiled achievement across the three cognitive domains, tested how the profile changed between cycles, and modelled each domain as a function of mathematics self-confidence, liking mathematics, home educational resources, and gender. All estimates used the five plausible values with jackknife (JK2) standard errors; domain contrasts are reported with both between-student (d_{av}) and paired (d_z) effect sizes, and student-level results were corroborated with two-level (student-within-school) models. Contrary to the reasoning-deficit narrative, reasoning was not the weakest domain in either cycle: in 2019 it exceeded applying and knowing, and by 2023 the domains converged, with knowing becoming the relative low point. The 2019–2023 gain was concentrated in the applying domain (+18.4 points, $p = .002$). Across all domains and both cycles, self-confidence ($\beta = .42-.45$) and home educational resources ($\beta = .33-.41$) were the dominant correlates, whereas liking mathematics contributed little once self-confidence was controlled (a suppression pattern), and gender differences were small. The predictor structure was near-identical across domains and stable across cycles; multilevel models further showed the home-resources association to be largely compositional (between schools). These findings reframe the Turkish “reasoning gap” and point to school composition and mathematics self-confidence as policy priorities.

Keywords:

TIMSS, mathematics cognitive domains, reasoning, mathematics self-confidence, Türkiye, large-scale assessment.



www.iejee.com
ISSN: 1307-9298

2026 Published by KURA Education & Publishing.
This is an open access article under the CC BY-NC-ND license. (<https://creativecommons.org/licenses/by/4.0/>)

Introduction

In an increasingly technological society, mathematical competence has become a core skill for all learners, equipping individuals to adapt and meet new challenges

as the demands of the world continue to shift (Philpot et al., 2021). Given this importance, mathematics achievement remains a central focus of education systems and is systematically monitored through international assessments. Among these, the Trends in International Mathematics and Science Study (TIMSS), conducted by the International Association for the Evaluation of Educational Achievement (IEA), has assessed the mathematics achievement of eighth-grade students every four years since 1995, providing nearly three decades of trend data that allow comparisons both over time and across countries (von Davier et al., 2024).

Mathematics is characterized not only by a body of concepts and procedures but also by distinctive ways of reasoning through which mathematical knowledge is constructed, justified, and extended (Ball & Bass, 2003). Consistent with this view, mathematical proficiency is widely regarded as multidimensional rather than a single body of content knowledge. Kilpatrick et al. (2001) conceptualize it as comprising interwoven strands of conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition, while Niss (2003) likewise frames the goal of mathematics education in terms of developing mathematical competencies rather than acquiring content knowledge alone. Reflecting this multidimensional perspective, TIMSS conceptualizes mathematics achievement along both content and cognitive dimensions.

These two dimensions are specified in detail in the TIMSS 2023 mathematics framework. The content dimension specifies the subject matter assessed (number, algebra, geometry, and data and probability), whereas the cognitive dimension specifies the thinking processes students draw on when working with that content. Within each content domain, items are developed to assess all three cognitive domains, so that number, algebra, geometry, and data items alike may require knowing, applying, or reasoning; the cognitive dimension therefore captures how students engage with mathematics, independent of the particular topic. It comprises three domains. Knowing refers to students' recall and understanding of mathematical facts, concepts, and procedures. Applying concerns their ability to use this knowledge to solve routine and real-world problems by selecting appropriate strategies and representing situations mathematically. Reasoning involves the logical, systematic thinking required to analyze complex relationships, make inferences, generalize, and justify solutions (Philpot et al., 2021). This design enables researchers to examine students' achievement profiles across knowing, applying, and reasoning, providing a more nuanced picture than overall scores alone. Although the content and cognitive dimensions are conceptually distinct, performance in each cognitive

domain necessarily draws on students' mathematical content knowledge. This organization parallels the revised taxonomy of Anderson and Krathwohl (2001), which likewise conceptualizes learning as the intersection of knowledge and cognitive processes. Accordingly, TIMSS reports not only overall mathematics achievement but also achievement in each cognitive domain separately (Yin & Foy, 2021), making it possible to identify cognitive achievement profiles that may be obscured by overall scores.

Beyond overall achievement levels, a substantial body of research using TIMSS has examined the student characteristics associated with achievement. Across countries and assessment cycles, mathematics self-confidence, home educational resources, liking mathematics, and gender have emerged as the most frequently investigated student-level variables and have consistently been associated with mathematics achievement (Akyüz, 2014; Lee & Stankov, 2018; Yıldırım, 2019).

Among these variables, mathematics self-confidence has consistently emerged as the strongest correlate of mathematics achievement. International syntheses and meta-analyses show that students with higher mathematics self-confidence achieve substantially higher mathematics scores across countries (Çiftçi & Yıldız, 2019; Lee & Stankov, 2018). Similar evidence has been reported for Türkiye. Using TIMSS 2011, 2015, and 2019 data, Coşkun and Karakaya Özyer (2023) found that mathematics self-confidence was the strongest student-level predictor of mathematics achievement across all three assessment cycles, whereas the unique effects of liking mathematics and valuing mathematics were relatively weak or inconsistent. Home educational resources likewise demonstrate a robust positive association with mathematics achievement in Türkiye (Coşkun & Karakaya Özyer, 2023; Ersan & Rodriguez, 2020), a finding that is broadly consistent with other analyses of Turkish TIMSS data (Koyuncu, 2021).

Gender has also received considerable attention in TIMSS research. However, large cross-national studies indicate that gender differences in mathematics achievement are generally small and inconsistent across countries, suggesting that affective and socioeconomic characteristics explain substantially more variation than gender alone (Else-Quest et al., 2010; Ghasemi & Burley, 2019; Reilly et al., 2019; Oberleiter et al., 2023). Recent evidence from Türkiye similarly suggests that gender differences vary across TIMSS cycles, with female students outperforming males in some cycles but not others (Coşkun & Karakaya Özyer, 2023). Collectively, these findings suggest that mathematics self-confidence and home educational resources are the most consistent correlates of mathematics achievement, whereas

the unique contributions of liking mathematics and gender appear to vary across contexts and assessment cycles. Although this body of research has substantially advanced the understanding of mathematics achievement in TIMSS, it has almost exclusively conceptualized achievement as a single composite outcome. Whether these well-established correlates operate uniformly across the knowing, applying, and reasoning domains that constitute mathematics achievement—or whether their relative importance shifts with the cognitive demands of the tasks—therefore remains an open empirical question.

Because the three TIMSS cognitive domains represent qualitatively different mathematical processes, such variation in the relative importance of these correlates is theoretically plausible. Yet despite this conceptual and practical importance, the way achievement is distributed across the cognitive domains has received comparatively little empirical attention. The broader body of research built on international large-scale assessments such as TIMSS has grown rapidly into a mature field (Hernández-Torrano & Courtney, 2021), yet in a mapping of its dominant research themes, achievement disaggregated by the knowing–applying–reasoning distinction does not emerge as a distinct topic. The field’s attention has instead centered on themes such as measurement and testing, educational policy, equity, and student motivation, with the most visible work focusing on cross-national differences in overall achievement (Hernández-Torrano & Courtney, 2021). The process-oriented cognitive dimension has likewise been comparatively neglected within more fine-grained work. Zhu (2023) notes that most research on mathematics learning trajectories concentrates on knowledge content rather than the cognitive domain. Where the cognitive domains are examined directly, analyses have frequently remained at the level of the assessment framework rather than learner performance. Choi et al. (2026), for instance, explicitly acknowledge that their document-based analysis “does not directly examine learner-level performance,” leaving empirical validation against actual achievement data as a task for future research.

Türkiye offers a particularly informative case for taking up this task. With a long record of participation at the eighth-grade level—in 1999, 2007, 2011, 2015, 2019, and 2023—it provides the sustained, learner-level trend data needed to examine cognitive-domain achievement both over time and against actual student performance rather than at the framework level alone. Its recent trajectory makes such an examination especially worthwhile. Across these cycles, Türkiye’s average mathematics scores have risen consistently. In TIMSS 2023, Turkish eighth-graders achieved a mathematics score of 509, well above the international average of 478 and ranking 13th among the 44 participating countries. Notably, this marks the first time Türkiye surpassed the TIMSS scale

centerpoint of 500 at this grade level, with an increase of 13 points over the 2019 cycle. Beneath this overall figure, however, the three cognitive domains reveal an uneven profile. Turkish eighth-graders scored 505 in knowing, 510 in applying, and 511 in reasoning (Ministry of National Education [MoNE], 2024). This profile is noteworthy because Türkiye is one of only eight countries where reasoning is the single highest-scoring cognitive domain. This ordering also runs counter to the difficulty hierarchy ordinarily expected across the domains, in which knowing is the easiest and reasoning the most demanding (Zhu, 2023).

Although the TIMSS 2023 Türkiye Report and the international results document this cognitive-domain profile, these findings remain largely descriptive. The international report does test each domain score against a country’s overall mathematics average, and on this test Türkiye’s knowing, applying, and reasoning scores do not differ significantly from its overall average at the .01 level (von Davier et al., 2024; Ministry of National Education [MoNE], 2024). Yet this leaves the questions most relevant to interpreting Türkiye’s profile unanswered. First, comparing each domain with the overall composite—which itself incorporates that domain—does not establish whether the three domains differ from one another; whether the descriptive ordering in which reasoning ranks highest reflects a genuine relative strength or merely sampling variation therefore remains untested. Second, the reports do not indicate whether the cognitive-domain profile changed significantly between the 2019 and 2023 cycles. Third, they do not show whether student-level characteristics—such as self-confidence, liking mathematics, gender, and home educational resources—are differentially associated with achievement across the three domains. The present study addresses these gaps by providing an inferential analysis of Turkish eighth-graders’ mathematics achievement across the knowing, applying, and reasoning domains using TIMSS 2019 and 2023 data.

The Present Study: Aims and Hypotheses

To this end, the TIMSS 2019 and 2023 Grade 8 data for Türkiye were analyzed in pursuit of three aims: (a) to profile mathematics achievement across the three cognitive domains and quantify their relative standing; (b) to test whether the domain profile changed between 2019 and 2023; and (c) to examine how mathematics self-confidence, liking mathematics, home educational resources, and gender are associated with achievement in each domain, and whether these associations differ across domains or cycles. Because the design is observational, associations are not interpreted causally.

Four hypotheses were advanced. H1 (relative standing): Reasoning achievement will differ from knowing and applying achievement within each cycle; given

competing expectations in the literature, the direction is tested two-sided. H2 (profile change): The cognitive-domain profile will change between 2019 and 2023. H3 (dominant correlates): Mathematics self-confidence and home educational resources will be the strongest positive correlates of achievement in every domain, whereas liking mathematics will show weak unique associations once self-confidence is controlled. H4 (structural invariance): The configuration of predictors will be largely invariant across the three cognitive domains.

Method

Data and Sample

TIMSS 2019 and 2023 Grade 8 assessments for Türkiye, obtained from the IEA international database were used in the analysis (Mullis et al., 2020). TIMSS uses a two-stage stratified cluster sample in which schools are selected first and intact classrooms within schools second. The Türkiye main samples comprised 4,077 students in 181 schools (2019) and 4,925 students in 141 schools (2023). Only the main (digital) samples were analyzed; the paper bridge and problem-solving and inquiry sub-samples were excluded.

Measures

Outcomes

Mathematics achievement in each cognitive domain—knowing (BSMKNO), applying (BSMAPP), and reasoning (BSMREA)—was represented by five plausible values reported on the TIMSS scale (international centerpoint = 500, $SD = 100$). Plausible values are multiple imputations of proficiency that, combined appropriately, yield unbiased population estimates and correct standard errors (Arikan et al., 2020; Braun & von Davier, 2017).

Predictors

Three IEA-constructed IRT context scales were used: mathematics self-confidence (Students Confident in Mathematics, BSBGSCM), liking mathematics (Students Like Learning Mathematics, BSBGSLM), and home educational resources (BSBGHER); higher scores indicate more of the named attribute. Gender was coded 1 = girl, 0 = boy (ITSEX). Students Confident in Mathematics comprised eight Likert items (e.g., “I usually do well in mathematics”; Cronbach’s $\alpha = .89$ in the Turkish 2023 sample), and Students Like Learning Mathematics comprised nine items (e.g., “I enjoy learning mathematics”; $\alpha = .93$). Home Educational Resources is an IEA-scaled composite index combining the number of books in the home, home study supports (own room and internet access), and parents’ highest education level, and is therefore reported as an index rather than an internal-consistency scale.

Analytic Strategy

Analyses were conducted in R 4.5.2 (R Core Team, 2025) using the five plausible values and the total student weight (TOTWGT). Standard errors were estimated with the jackknife repeated-replication procedure (JK2) using the 75 sampling zones and the jackknife replicate indicator, and plausible-value estimates were combined by Rubin’s (1987) rules (Foy & LaRoche, 2020); this implementation reproduced the published Türkiye mathematics means and standard errors, confirming its accuracy. For each cycle we estimated weighted domain means and standardized pairwise domain differences. Two effect sizes are reported for the domain contrasts: d_{av} , standardized by the pooled between-student standard deviation, and—because the domains are strongly correlated within students—the paired effect size d_z , standardized by the standard deviation of the within-student difference scores (Cohen, 1988). Cross-cycle contrasts were tested with independent-samples z tests based on the two cycles’ jackknifed standard errors. Predictors of each domain were estimated with weighted linear regressions (predictors standardized), fitted separately by domain and cycle and pooled across the five plausible values; coefficients are interpreted as associations. Missing data on predictors were limited (< 3.5% per variable) and handled by listwise deletion.

Results

Following the analytic approach described in Section 2.3, all estimates are weighted (TOTWGT) and pool the five plausible values, and are reported separately for TIMSS 2019 ($N = 4,077$) and 2023 ($N = 4,925$). The three cognitive domains—knowing, applying, and reasoning—are examined first as an achievement profile (weighted domain means and standardized pairwise differences) and then as outcomes in parallel regression models predicted by mathematics self-confidence, liking mathematics, home educational resources, and gender.

Descriptive Statistics and the Cognitive-Domain Profile

Table 1 reports weighted descriptive statistics for all study variables in each cycle, and Figure 1 displays the cognitive-domain achievement profile. In 2019, reasoning ($M = 503.9$) stood above both knowing ($M = 494.1$) and applying ($M = 491.3$); by 2023 all three domains had risen and clustered more closely together (knowing = 504.7, applying = 509.8, reasoning = 511.4), each at or above the TIMSS scale centerpoint of 500. Thus, contrary to the common expectation of a reasoning shortfall, Grade 8 students in Türkiye did not underperform in reasoning relative to the more routine domains in either cycle; instead, the domain that lagged shifted over time.

Table 1

Weighted Descriptive Statistics for Study Variables by TIMSS Cycle

Variable	M 2019	SD 2019	M 2023	SD 2023
Reasoning achievement	503.9	110.6	511.4	111.8
Applying achievement	491.3	103.4	509.8	111.4
Knowing achievement	494.1	126.0	504.7	120.6
Self-confidence (BSBGSCM)	9.84	2.37	9.65	2.34
Liking mathematics (BSBGSLM)	10.32	1.93	10.22	1.97
Home resources (BSBGHER)	9.47	1.83	10.23	1.66
Female (%)	50.2	—	51.4	—
N	4,077		4,925	

Note. Achievement domains are on the TIMSS reporting scale (international centerpoint = 500); context measures are IEA IRT scales. All estimates are weighted (TOTWGT). 2019 N = 4,077; 2023 N = 4,925.

Table 2

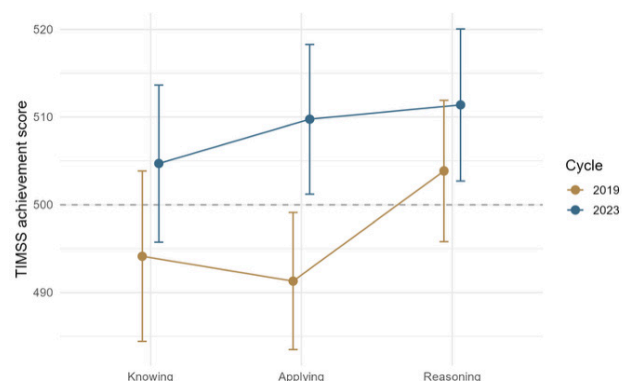
Weighted Bivariate Correlations Among Study Variables

Variable	1	2	3	4	5	6	7
1. Reasoning	—	.96	.96	.44	.25	.47	.06
2. Knowing	.95	—	.96	.47	.27	.48	.07
3. Applying	.97	.96	—	.45	.25	.47	.03
4. Self-confidence	.49	.50	.48	—	.70	.14	-.03
5. Liking mathematics	.31	.31	.30	.67	—	-.01	.02
6. Home resources	.43	.43	.43	.22	.08	—	.05
7. Female	-.01	.02	-.01	-.07	-.04	.00	—

Note. Weighted correlations averaged across five plausible values for achievement domains. Values above the diagonal are TIMSS 2019; below the diagonal are TIMSS 2023. Female coded 1 = girl.

Figure 1

Mathematics Cognitive-Domain Achievement Profile, Türkiye Grade 8



Note. Weighted means with 95% confidence intervals (five plausible values, JK2 jackknife). The dashed line marks the TIMSS scale centerpoint (500).

Pairwise Domain Differences and Cross-Cycle Change

In 2019, reasoning significantly exceeded applying by 12.5 score points ($p < .001$, $d_{av} = 0.12$) and knowing by 9.7 points ($p < .001$, $d_{av} = 0.08$), whereas applying and knowing were practically indistinguishable. By 2023 the reasoning–applying difference had shrunk to

a non-significant 1.6 points ($p = .17$, $d_{av} = 0.01$), and knowing had become the lowest-performing domain: applying exceeded knowing by 5.1 points ($p < .001$) and reasoning exceeded knowing by 6.7 points ($p < .001$). Across cycles (Table 3), the overall improvement was driven primarily by applying, which rose 18.4 points ($p = .002$, $d = 0.17$); the gains in knowing (+10.6) and reasoning (+7.5) were positive but not statistically significant. Consistent with this, the reasoning advantage over applying narrowed significantly ($\Delta = -10.9$, $p < .001$), and applying moved from slightly below to significantly above knowing ($\Delta = +7.9$, $p < .001$; Table 4). All within-cycle domain differences were small in effect-size terms ($|d_{av}| \leq 0.12$), indicating that the domains are close in absolute standing even where differences are statistically reliable. Because the domains were highly correlated within students ($r_s \approx .96$), paired effect sizes exceeded the between-student d_{av} ; for example, the 2019 reasoning-over-applying advantage corresponded to $d_{av} = 0.12$ but $d_z = 0.42$, declining to $d_z = 0.06$ by 2023 (Table 3).

Table 3

Cognitive-Domain Achievement and Cross-Cycle Change (2019 to 2023)

Domain	M 2019 (SE)	M 2023 (SE)	Δ	z	p	d
Knowing	494.1 (4.95)	504.7 (4.57)	+10.6	1.57	.117	0.086
Applying	491.3 (3.99)	509.8 (4.35)	+18.4	3.12	.002	0.172
Reasoning	503.9 (4.11)	511.4 (4.42)	+7.5	1.25	.212	0.068

Note. Weighted means with jackknife standard errors in parentheses. d = Cohen's d for the cross-cycle change. Cross-cycle tests are independent-samples z tests.

Table 4

Pairwise Cognitive-Domain Differences by Cycle and Their Cross-Cycle Change

Contrast	2019 diff (d_{av}/d_z)	2023 diff (d_{av}/d_z)	Δ gap	z	p
Reasoning - Applying	+12.5 (0.117/0.421)***	+1.6 (0.015/0.056)	-10.9	-5.54	< .001
Reasoning - Knowing	+9.7 (0.082/0.255)***	+6.7 (0.058/0.186)***	-3.0	-1.31	.189
Applying - Knowing	-2.8 (-0.025/-0.072)*	+5.1 (0.044/0.157)***	+7.9	4.61	< .001

Note. Positive values indicate the first domain exceeds the second. d = paired Cohen's d . * $p < .05$. ** $p < .01$. *** $p < .001$.

Predictors of Cognitive-Domain Achievement

The three cognitive domains were strongly intercorrelated ($r_s = .95-.97$; Table 2). Among the context measures, mathematics self-confidence showed the strongest bivariate associations with achievement ($r_s = .44-.50$), followed by home educational resources ($r_s = .43-.48$) and liking mathematics ($r_s = .25-.31$), whereas gender was essentially unrelated to achievement at the bivariate level ($|r_s| \leq .07$). Self-confidence and liking mathematics were themselves highly correlated ($r = .70$ in 2019, $.67$ in 2023), which anticipates the suppression pattern observed in the multivariable models.

Table 5 and Figure 2 present the standardized regression coefficients. Across all three domains and both cycles, mathematics self-confidence was the strongest predictor ($\beta = .42-.45$, $p < .001$), followed by home educational resources ($\beta = .33-.41$, $p < .001$). Once self-confidence was held constant, liking mathematics carried a negligible or slightly negative unique association with achievement (β from $-.05$ to $.00$), consistent with its high overlap with self-confidence. Girls scored marginally higher than boys, an association that was clearest for knowing and reasoning in 2019 ($\beta = .11-.13$) and attenuated toward zero by 2023. The models accounted for 35–40% of the variance in domain achievement ($R^2 = .35-.40$). Notably, the predictor structure was almost identical across the three domains: no domain was distinguished by a different leading predictor, and self-confidence and home resources dominated knowing, applying, and reasoning alike.

Table 5

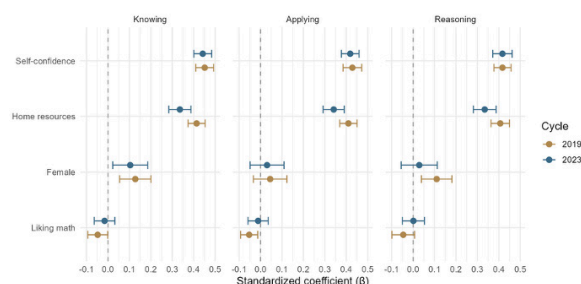
Standardized Regression Coefficients Predicting Cognitive-Domain Achievement by Cycle

Domain	Predictor	β 2019	β 2023	$\Delta\beta$	p ($\Delta\beta$)
Knowing	Self-confidence	.45***	.44***	-0.01	.771
	Home resources	.41***	.34***	-0.08	.019
	Female	.13***	.10*	-0.02	.670
	Liking mathematics	-.05*	-.02	+0.03	.347
Applying	Self-confidence	.43***	.42***	-0.01	.735
	Home resources	.41***	.34***	-0.07	.034
	Female	.04	.03	-0.01	.799
	Liking mathematics	-.05**	-.01	+0.04	.190
Reasoning	Self-confidence	.42***	.42***	+0.00	.995
	Home resources	.41***	.33***	-0.07	.037
	Female	.11**	.03	-0.08	.149
	Liking mathematics	-.05	.00	+0.05	.211

Note. Standardized coefficients from separate PV-pooled, jackknifed regressions per domain and cycle; all predictors entered simultaneously. $\Delta\beta = \beta_{2023} - \beta_{2019}$ (independent-samples test). Model R^2 (2019/2023): Knowing = 0.40/0.36; Applying = 0.37/0.35; Reasoning = 0.36/0.35. * $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 2

Predictors of Cognitive-Domain Achievement by Domain and Cycle, Türkiye Grade 8



Note. Standardized coefficients (β) with 95% confidence intervals from separate PV-pooled, jackknifed models per domain. Cycles are offset vertically for legibility.

Cross-cycle comparison of the coefficients revealed only one systematic change: the standardized effect of home educational resources declined modestly but significantly across all three domains ($\Delta\beta \approx -.07$ to $-.08$, $ps < .05$), while the coefficients for self-confidence, gender, and liking mathematics were stable (all $ps > .14$). Accordingly, the relative shortfall in knowing that emerged by 2023 cannot be attributed to a domain-specific configuration of student-level predictors; the same affective and resource factors predicted all three cognitive domains to a very similar degree in both cycles.

Multilevel robustness. To account for the nesting of students within schools, we re-estimated the predictor models as two-level random-intercept models. Between-school variation was substantial (ICC = .36 in 2019 and .33 in 2023), yet the central findings were unchanged: mathematics self-confidence remained the strongest predictor of overall achievement ($\beta = .43$ and $.40$), the gender coefficient was stable, and liking mathematics remained negligible. The one coefficient sensitive to the multilevel specification was home educational resources, which attenuated from $\beta = .42$ to $.21$ in 2019 and from $.35$ to $.17$ in 2023. A contextual model clarified this pattern: the within-school (individual) coefficient was small ($\beta = .17$ and $.15$), whereas the school-mean coefficient was large ($\beta = .36$ and $.33$), indicating that the home-resources association operates substantially through between-school composition rather than at the individual level.

Discussion

This study profiled eighth-grade mathematics achievement in Türkiye across the knowing, applying, and reasoning domains in two TIMSS cycles and examined how affective and resource factors relate to each domain. Four expectations guided the analysis. Consistent with H1, reasoning differed from the more routine domains within each cycle, though the differences were small; consistent with H2, the cognitive-domain profile changed between 2019 and 2023 rather than rising uniformly; consistent with H3, mathematics self-confidence and home educational resources were the dominant correlates in every domain while liking mathematics added little unique variance; and consistent with H4, the configuration of predictors was essentially invariant across the three domains. We develop these findings in turn, giving particular attention to what they can and cannot support given the observational design and the measurement properties of cognitive subscores.

No reasoning deficit. Contrary to a widespread assumption, reasoning was not the weakest domain in either cycle. In 2019 it was the relative strength, exceeding applying by 12.5 score points ($d_{av} = 0.12$) and knowing by 9.7 points ($d_{av} = 0.08$); by 2023 the three domains had converged, with reasoning

and applying statistically indistinguishable and knowing the relative low point. Two features of this result warrant emphasis. First, although several within-cycle differences were statistically reliable, all were small in effect-size terms ($|d| \leq 0.12$), so the domains are better described as close in standing than as sharply differentiated—a reading reinforced by their very high intercorrelations ($r_s = .95-.97$). Statistical significance here reflects the large sample and the strong within-student correlation among domains as much as any substantive gap. Second, the ordering in which reasoning is not a weakness is not new for Türkiye. The Ministry's own TIMSS 2019 reporting noted comparatively strong reasoning performance (MoNE, 2020), and the present estimates place that observation on an inferential footing. The pattern qualifies, without wholly overturning, the familiar “strong on procedures, weak on reasoning” characterization: analyses of the intended Turkish middle-school curriculum indicate that reasoning demands are nominally well represented, particularly at Grade 8 (Bulut & Taşpınar Şener, 2023), whereas studies of textbooks and classroom tasks document a predominance of low-demand, procedure-oriented work (Polat & Dede, 2023; Horzum & Duran, 2024). The proceduralist narrative may therefore describe the implemented curriculum more accurately than the intended one. Two caveats temper the reframing. The convergence observed here is not unique to Türkiye. Internationally, reasoning shifted from a common relative strength in 2019 to a common relative weakness in 2023 (Mullis et al., 2020; von Davier et al., 2024), so Türkiye's trajectory partly tracks a global movement. And because TIMSS cognitive subscales rest on item subsets and are less robust than the overall score, the profile—consistent with the domain-profiling approach adopted elsewhere (Mokgwathi, 2026)—should be read as indicative rather than definitive.

A shifting profile, led by applying. The 2019–2023 improvement was concentrated in the applying domain, which rose 18.4 points ($d = 0.17$) and was the only domain whose cross-cycle gain reached significance; the gains in knowing (+10.6) and reasoning (+7.5) were positive but not reliable. Correspondingly, applying moved from slightly below to significantly above knowing, and the reasoning advantage over applying narrowed. The profile thus changed in shape rather than simply shifting upward. This redistribution warrants one clarification: because knowing achievement still rose numerically (+10.6 points), its emergence as the relative low point reflects a slower rate of improvement, not a decline in factual or procedural competence—reading it as deterioration would overstate what the data support. Three readings compete to explain the change. The most plausible is the cumulative influence of the skill-based assessment environment that expanded

following the introduction of Türkiye's high-stakes transition examination, in which mathematics items increasingly emphasized contextual interpretation, inference, and the functional use of knowledge over routine recall (Ekinçi & Bal, 2019; Erden, 2020; Uzun & Ağaç, 2023; Bozkuş & Özgeldi, 2024). This is better characterized as a sustained assessment and instructional emphasis than as a formal curricular reform, since Türkiye's most recent curriculum revision postdates the 2023 administration. A pandemic reading is more equivocal. The 2023 cohort's schooling was disrupted by closures associated with reduced participation and mathematics learning loss in Türkiye (Haser et al., 2022) and with substantial but heterogeneous losses internationally (Gajderowicz et al., 2026), so the applying gains might signal selective recovery—but the relative weakness in knowing cannot be attributed specifically to the pandemic with these data. An instrumentation reading is weaker than it first appears: the Grade 8 cognitive-domain target weights were unchanged across cycles (Mullis & Martin, 2017), and because Türkiye had already administered the digital assessment in 2019 (Yalçın, 2022), a simple paper-to-digital mode effect is unlikely (Fishbein et al., 2018). Even so, TIMSS 2023 strengthened its use of contextualized, interactive items and a group-adaptive design (von Davier et al., 2024), so differences in item context and composition remain a residual consideration alongside the subscale-reliability caveat.

Self-confidence and home resources dominate; liking adds little. Across every domain and both cycles, mathematics self-confidence was the strongest correlate of achievement ($\beta = .42-.45$), followed by home educational resources ($\beta = .33-.41$); once self-confidence was controlled, liking mathematics carried a negligible or slightly negative unique association (β from $-.05$ to $.00$). Because self-confidence and liking were themselves highly correlated ($r = .70$ in 2019, $.67$ in 2023), this pattern is consistent with suppression rather than with enjoyment being irrelevant to achievement. The dominance of self-confidence and home resources aligns with prior Turkish TIMSS work (Ersan & Rodriguez, 2020; Coşkun & Karakaya Özyer, 2023) and with meta-analytic evidence for a robust confidence–achievement association (Çiftçi & Yıldız, 2019). Interpreting it, however, requires care about what TIMSS measures. The confidence scale indexes a competence-related self-belief closer to academic self-concept—normative and comparatively stable—than to task-specific self-efficacy (Bong & Skaalvik, 2003), and self-concept and achievement are reciprocally related over time (Marsh & Craven, 2006; Marsh & Martin, 2011). Cross-sectional dominance therefore cannot establish causal direction. Nor should the weak adjusted coefficient for liking be read as showing enjoyment to be educationally inert: situated expectancy–value theory assigns competence beliefs

a more proximal role in performance while enjoyment remains consequential for engagement, persistence, and later choices (Eccles & Wigfield, 2020), and the attenuation of a liking effect once confidence is entered is a recurring TIMSS pattern (Østbø & Zachrisson, 2022; Ivanova & Michaelides, 2022). On balance, the evidence points—tentatively—toward prioritizing competence-related confidence through mastery experiences and strategy-focused feedback while continuing to cultivate interest (Usher & Pajares, 2009). Because the design cannot show confidence to be causally prior, this is best read as a direction for intervention research rather than a warranted prescription.

The same factors predict all domains. The predictor configuration was essentially invariant across knowing, applying, and reasoning: no domain was distinguished by a different leading correlate, and the models accounted for similar variance throughout ($R^2 = .35-.40$). A comparably domain-general pattern has been reported for Turkish fourth-graders in TIMSS 2019 (Ayva Yörü et al., 2023). This invariance is informative, but not in the way it may first appear: a predictor set that does not vary across domains cannot, by construction, account for variation between them. Student-level affect and resources are therefore silent on the shortfall in knowing that emerged by 2023—not because they were tested and failed, but because they are the wrong grain of analysis for a domain-specific gap. That question belongs more properly to students' opportunities to learn. At the level of the intended curriculum, knowing appears under-weighted relative to the TIMSS Grade 8 targets, while reasoning is nominally well represented but concentrated in lower-order analytic outcomes, with generalizing and justifying scarcely present (Bulut & Taşpınar Şener, 2023); central examinations have moved in a similar direction, with knowledge-level items declining across cycles (Bostan-Sarioğlu et al., 2021; Ekinci & Bal, 2019; Uzun & Ağa, 2023). A relative weakness in knowing is thus less an anomaly to be explained away than a plausible consequence of what students are given the opportunity to learn (Schmidt et al., 1997). Evidence on the enacted curriculum is less consistent: textbook and classroom studies document substantial reliance on low-demand procedural tasks and teacher-led instruction (Şensoy & Kılıç, 2021; Polat & Dede, 2023; Horzum & Duran, 2024), teachers report gaps between skill-based assessment demands and available curricular guidance (Erden, 2020; Kertil et al., 2021), and teacher quality and working conditions account for meaningful between-school variation in achievement (Sezer & Çakan, 2022). This account remains provisional; TIMSS itself cautions that content coverage is only one component of opportunity to learn and that enacted curricula frequently diverge from intended ones (Bakan Kalaycioglu et al., 2026). Establishing it would require linking domain-level

performance to teacher-reported coverage rather than inferring it from curriculum documents alone.

Small, diminishing gender gaps. Gender differences were small and, where detectable, favored girls in knowing and reasoning in 2019 ($\beta = .11-.13$), attenuating toward zero by 2023. This limited differentiation is consistent with earlier Turkish analyses in which gender did not significantly predict overall mathematics achievement net of student and school characteristics (Coşkun & Karadağ, 2023; Şahin et al., 2025) and with the international finding of broad parity at Grade 8 (Mullis et al., 2020); cross-national work likewise indicates that mathematics gender gaps are small and variable in magnitude and direction (Else-Quest et al., 2010; Meinck & Brese, 2019). The attenuation admits a substantive reading rather than mere regression to the mean: pandemic-related learning losses were not gender-neutral, falling disproportionately on girls in international analyses (Gajderowicz et al., 2026), so the narrowing may reflect an erosion of girls' prior advantage rather than gains among boys—a distinction these data cannot adjudicate but which future cycles should. Achievement parity should not, in any case, be equated with affective parity: TIMSS-based studies find that girls often report less positive mathematics self-beliefs even where their achievement matches or exceeds that of boys (Ghasemi & Burley, 2019; Mejía-Rodríguez et al., 2021), a gap with potential implications for later participation in advanced mathematics and STEM.

Achievement inequality is largely a between-school phenomenon. The multilevel models added an important qualification. Although home educational resources predicted achievement strongly in single-level models, most of this association operated between schools rather than within them (within-school $\beta \approx .15-.17$ vs. school-mean $\beta \approx .33-.36$), against a backdrop of substantial between-school variance ($ICC \approx .33-.36$). Students' achievement was thus bound up not only with their own resources but with the resource composition of the school they attended. This is consistent with earlier Turkish TIMSS evidence that socioeconomic status predicts achievement more strongly at the school than the student level (Ersan & Rodriguez, 2020) and that a substantial share of eighth-grade achievement variance lies between schools (Suna & Özer, 2021), as well as with national and international findings that students of differing backgrounds are unevenly distributed across school types closely tied to composition, resources, and learning environments (Alacacı & Erbaş, 2010; Özdemir, 2016; Suna et al., 2020; Özer & Suna, 2023). The mechanism, however, requires care: because these are Grade 8 students who have not yet entered upper-secondary tracks, the finding cannot be read as evidence that high-school tracking produced the observed variance. Rather, it situates achievement

within a system in which between-school differences emerge before upper-secondary placement and are subsequently reinforced by selective transitions (Bölükbaş & Gür, 2020; Niehues et al., 2019). The policy implication is that narrowing achievement gaps is unlikely to be achieved through student-level interventions alone and may require attention to school composition and the distribution of resources across schools. Mathematics self-confidence, by contrast, remained a robust individual-level correlate and is tempting to read as a malleable within-school lever; that reading warrants caution, since the confidence scale indexes academic self-concept, which is itself compositionally shaped—equally able students tending to report lower self-concept in higher-achieving settings (Marsh & Craven, 2006; Bong & Skaalvik, 2003).

Limitations and future directions

Several constraints qualify these findings. The design is observational and cross-sectional within each cycle; although single-level estimates are complemented by multilevel models, no causal claims follow, and this matters most for the plausibly bidirectional confidence–achievement association (Marsh & Craven, 2006). Measurement invariance of the context scales was assumed rather than formally tested across cycles and gender, so apparent change in an affective predictor could partly reflect a shift in measurement rather than in the construct (Atılğan & Deniz, 2023). The cycles also differ instrumentally: TIMSS 2023 completed the transition to a fully digital, group-adaptive assessment with new interactive formats (von Davier et al., 2024); although Türkiye's prior digital administration makes a simple mode effect unlikely (Yalçın, 2022), differences in item context remain a possible confound for profile comparison. The profile rests on two cycles only, so the convergence reported here cannot yet be distinguished from cycle-specific fluctuation; and the 2023 cohort's schooling was disrupted unevenly by the pandemic (Gajderowicz et al., 2026), so cohort effects are not fully separable from trend. Finally, cognitive-domain subscores rest on smaller item subsets and are measured less precisely than the overall score, so the small differences among knowing, applying, and reasoning should be interpreted alongside the reported effect sizes and confidence intervals rather than as substantive findings in their own right. These constraints converge on a common remedy: longitudinal or cross-lagged designs to establish temporal ordering, formal invariance testing before cross-cycle and cross-group comparison, and replication on mode-invariant trend items across additional cycles.

Conclusion

Disaggregating Turkish eighth-graders' mathematics achievement by cognitive domain yields a picture

that overall scores obscure: reasoning is not a national weakness, the recent improvement was led by applying while knowing lagged, and the same affective and resource factors predict all three domains alike. Because those student-level factors cannot, by construction, explain a domain-specific gap, the emerging shortfall in knowing points instead toward opportunity-to-learn, and the strong between-school patterning of resources locates much of the inequality at the level of schools rather than individuals.

More broadly, these findings illustrate the value of examining mathematics achievement through its cognitive dimensions rather than overall scores alone. Overall achievement remains an indispensable indicator for monitoring educational systems, but a domain-level lens reveals patterns—an applying-led rise, a knowing shortfall unexplained by student-level factors, and inequality concentrated between schools—those composite scores leave hidden. Read with due caution about the observational design and the reliability of subscores, the results suggest that both curricular alignment and the equitable distribution of school resources deserve closer attention than overall achievement trends alone would imply, and that pairing cognitive-domain achievement with classroom, teacher, and curriculum data is a priority for future research.

References

- Akyüz, G. (2014). The effects of student and school factors on mathematics achievement in TIMSS 2011. *Education and Science*, 39(172), 150–162. <https://doi.org/10.15390/ES.2014.1255>
- Alacacı, C., & Erbaş, A. K. (2010). Unpacking the inequality among Turkish schools: Findings from PISA 2006. *International Journal of Educational Development*, 30(2), 182–192. <https://doi.org/10.1016/j.ijedudev.2009.03.006>
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed.). Longman.
- Arıkan, S., Özer, F., Şeker, V., & Ertaş, G. (2020). The importance of sample weights and plausible values in large-scale assessments. *Journal of Measurement and Evaluation in Education and Psychology*, 11(1), 43–60. <https://doi.org/10.21031/epod.602765>
- Atılğan, M., & Deniz, K. Z. (2023). Investigation of the measurement invariance of affective characteristics related to TIMSS 2019 mathematics achievement by gender. *Journal of Measurement and Evaluation in Education and Psychology*, 14(3), 185–199. <https://doi.org/10.21031/epod.1221365>

- Ayva Yörü, F. G., Sezer Başaran, E., & Çakan, M. (2023). Okul, ev ve öğrenci özelliklerinin 4. sınıf öğrencilerinin TIMSS 2019 matematik bilişsel becerilerindeki rolü [The role of school, home and student characteristics in TIMSS 2019 mathematics cognitive skills of 4th grade students]. *Mehmet Akif Ersoy University Journal of Education Faculty*, 67, 197–227. <https://doi.org/10.21764/mauefd.1142768>
- Bakan Kalaycioglu, D., Reynolds, K. A., Fishbein, B., Gonzalez, E., & von Davier, M. (2026). *Intended, implemented, and attained mathematics and science curricula at the eighth grade: Findings from TIMSS 2023*. TIMSS & PIRLS International Study Center, Boston College. <https://doi.org/10.6017/lse.tpisc.timss.wf9486>
- Ball, D. L., & Bass, H. (2003). Making mathematics reasonable in school. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 27–44). National Council of Teachers of Mathematics.
- Bölükbaş, S., & Gür, B. S. (2020). Tracking and inequality: The results from Turkey. *International Journal of Educational Development*, 78, Article 102262. <https://doi.org/10.1016/j.ijedudev.2020.102262>
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1), 1–40. <https://doi.org/10.1023/A:1021302408382>
- Bostan-Sarioğlu, A., Dolu, G., & Sevim, N. (2021). Sekizinci sınıf merkezi sınavlardaki fen sorularının TIMSS-2019 bilişsel alanlarına göre analizi [Analysis of science questions in eighth grade central exams according to cognitive fields of TIMSS-2019]. *e-Kafkas Journal of Educational Research*, 8(3), 514–533. <https://doi.org/10.30900/kafkasegt.973021>
- Bozkuş, Ç., & Özgeldi, M. (2024). Analysis of middle school mathematics teachers' evaluations of skill-based questions within the framework of mathematical competencies. *International Journal of Turkish Education Sciences*, 12(1), 535–583. <https://doi.org/10.46778/goputeb.1425323>
- Braun, H., & von Davier, M. (2017). The use of test scores from large-scale assessment surveys: Psychometric and statistical considerations. *Large-scale Assessments in Education*, 5(1), Article 17. <https://doi.org/10.1186/s40536-017-0050-x>
- Bulut, A. S., & Taşpınar Şener, Z. (2023). Analysis of secondary school mathematics curriculum learning outcomes by TIMSS-2019 cognitive domain skills. *Türk Akademik Yayınlar Dergisi (TAY Journal)*, 7(1), 303–328. <https://doi.org/10.29329/tayjournal.2023.53714>
- Choi, K. M., Mun, J., & Hand, B. (2026). Bridging reasoning in mathematics and science: A document analysis of applying and reasoning in the TIMSS assessment framework and the case for utilization of resources. *Educational Psychology Review*, 38, Article 22. <https://doi.org/10.1007/s10648-025-10115-2>
- Çiftçi, Ş. K., & Yıldız, P. (2019). The effect of self-confidence on mathematics achievement: The meta-analysis of Trends in International Mathematics and Science Study (TIMSS). *International Journal of Instruction*, 12(2), 683–694. <https://doi.org/10.29333/iji.2019.12243a>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- 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. <https://doi.org/10.33902/JPR.202318875>
- Coşkun, B., & Karakaya Özyer, K. (2023). The effect of student characteristics and socioeconomic status on mathematics achievement in Türkiye: Insights from TIMSS 2011-2019. *International Journal of Assessment Tools in Education*, 10(3), 454–481. <https://doi.org/10.21449/ijate.1272517>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61, Article 101859. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Ekinci, O., & Bal, A. P. (2019). 2018 yılı Liseye Geçiş Sınavı (LGS) matematik sorularının öğrenme alanları ve Yenilenmiş Bloom Taksonomisi bağlamında değerlendirilmesi [Evaluation of High School Entrance Exam (LGS) 2018in terms of mathematics learning field and Revised Bloom Taxonomy]. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi*, 7(3), 9–18. <https://doi.org/10.18506/anemon.462717>

- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136(1), 103–127. <https://doi.org/10.1037/a0018053>
- Erden, B. (2020). Türkçe, matematik ve fen bilimleri dersi beceri temelli sorularına ilişkin öğretmen görüşleri [Teachers' views related to skill-based questions in Turkish, mathematics and science lessons]. *Academia Journal of Educational Research*, 5(2), 270–292. <https://izlik.org/JA67NL34CY>
- Ersan, Ö., & Rodriguez, M. C. (2020). Socioeconomic status and beyond: A multilevel analysis of TIMSS mathematics achievement given student and school context in Turkey. *Large-scale Assessments in Education*, 8, Article 15. <https://doi.org/10.1186/s40536-020-00093-y>
- Fishbein, B., Martin, M. O., Mullis, I. V. S., & Foy, P. (2018). The TIMSS 2019 item equivalence study: Examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education*, 6, Article 11. <https://doi.org/10.1186/s40536-018-0064-z>
- Foy, P., & LaRoche, S. (2020). Estimating standard errors in the TIMSS 2019 results. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report*. TIMSS & PIRLS International Study Center, Boston College
- Gajderowicz, T., Jakubowski, M., Kennedy, A. I., Kjeldsen, C. C., Patrinos, H. A., & Strietholt, R. (2026). The learning crisis: Three years after COVID-19. *Learning and Instruction*, 104, Article 102373. <https://doi.org/10.1016/j.learninstruc.2026.102373>
- Ghasemi, E., & Burley, H. (2019). Gender, affect, and math: A cross-national meta-analysis of Trends in International Mathematics and Science Study 2015 outcomes. *Large-scale Assessments in Education*, 7, Article 10. <https://doi.org/10.1186/s40536-019-0078-1>
- Haser, Ç., Doğan, O., & Kurt Erhan, G. (2022). Tracing students' mathematics learning loss during school closures in teachers' self-reported practices. *International Journal of Educational Development*, 88, Article 102536. <https://doi.org/10.1016/j.ijedudev.2021.102536>
- Hernández-Torrano, D., & Courtney, M. G. R. (2021). Modern international large-scale assessment in education: An integrative review and mapping of the literature. *Large-scale Assessments in Education*, 9, Article 17. <https://doi.org/10.1186/s40536-021-00109-1>
- Horzum, T., & Duran, A. A. (2024). An examination of the cognitive demand levels reflected by mathematics tasks in the middle grades mathematics textbooks. *International Journal of Educational Spectrum*, 6(2), 294–321. <https://doi.org/10.47806/ijesacademic.1522050>
- Ivanova, M., & Michaelides, M. P. (2022). Motivational components in TIMSS 2015 and their effects on engaging teaching practices and mathematics performance. *Studies in Educational Evaluation*, 74, Article 101173. <https://doi.org/10.1016/j.stueduc.2022.101173>
- Kertil, M., Gülbağcı Dede, H., & Ulusoy, E. G. (2021). Skill-based mathematics questions: What do middle school mathematics teachers think about and how do they implement them? *Turkish Journal of Computer and Mathematics Education*, 12(1), 151–186. <https://doi.org/10.16949/turkbilmat.774651>
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. National Academy Press.
- Koyuncu, İ. (2021). TIMSS international benchmarks of eighth graders in mathematics: A correspondence analysis study. *International Electronic Journal of Elementary Education*, 14(2), 179–194. <https://doi.org/10.26822/iejee.2022.237>
- Lee, J., & Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences*, 65, 50–64. <https://doi.org/10.1016/j.lindif.2018.05.009>
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1(2), 133–163. <https://doi.org/10.1111/j.1745-6916.2006.00010.x>
- Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, 81(1), 59–77. <https://doi.org/10.1348/000709910X503501>
- Meinck, S., & Brese, F. (2019). Trends in gender gaps: Using 20 years of evidence from TIMSS. *Large-scale Assessments in Education*, 7, Article 8. <https://doi.org/10.1186/s40536-019-0076-3>

- Mejía-Rodríguez, A. M., Luyten, H., & Meelissen, M. R. M. (2021). Gender differences in mathematics self-concept across the world: An exploration of student and parent data of TIMSS 2015. *International Journal of Science and Mathematics Education*, 19(6), 1229–1250. <https://doi.org/10.1007/s10763-020-10100-x>
- Ministry of National Education. (2020). *TIMSS 2019 Türkiye preliminary report* [TIMSS 2019 Türkiye ön raporu]. Millî Eğitim Bakanlığı. https://www.meb.gov.tr/meb_iys_dosyalar/2020_12/10173505_No15_TIMSS_2019_Turkiye_On_Raporu_Guncel.pdf
- Ministry of National Education. (2024). *TIMSS 2023 Türkiye report* [TIMSS 2023 Türkiye raporu]. Directorate General for Measurement, Evaluation and Examination Services. <https://odsgm.meb.gov.tr/www/raporlar/kategori/96>
- Mokgwathi, M. S. (2026). Closing the gap in early mathematics: Domain and cognitive insights from TIMSS 2023 in South Africa and Singapore. *F1000Research*, 14, Article 1209. <https://doi.org/10.12688/f1000research.172015.3>
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., & von Davier, M. (Eds.). (2021). *TIMSS 2023 assessment frameworks*. TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. TIMSS & PIRLS International Study Center, Boston College.
- Niehues, W., Kisbu-Sakarya, Y., & Selçuk, B. (2019). Implications of between-school tracking for Turkish students. *Turkish Journal of Education*, 8(3), 196–216. <https://doi.org/10.19128/turje.453383>
- Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM project. In A. Gagatsis & S. Papastavridis (Eds.), *Proceedings of the 3rd Mediterranean Conference on Mathematical Education* (pp. 115–124). Hellenic Mathematical Society.
- Oberleiter, S., Fries, J., Schock, L. S., Steininger, B., & Pietschnig, J. (2023). Predicting cross-national sex differences in large-scale assessments of students' reading literacy, mathematics, and science achievement: Evidence from PIRLS and TIMSS. *Intelligence*, 100, Article 101784. <https://doi.org/10.1016/j.intell.2023.101784>
- Østbø, I. U., & Zachrisson, H. D. (2022). Student motivation and parental attitude as mediators for SES effects on mathematics achievement: Evidence from Norway in TIMSS 2015. *Scandinavian Journal of Educational Research*, 66(5), 808–823. <https://doi.org/10.1080/00313831.2021.1939138>
- Özdemir, C. (2016). Equity in the Turkish education system: A multilevel analysis of social background influences on the mathematics performance of 15-year-old students. *European Educational Research Journal*, 15(2), 193–217. <https://doi.org/10.1177/1474904115627159>
- Özer, M., & Suna, E. (2023). The relationship between school socioeconomic composition and academic achievement in Türkiye. *Journal of Economy Culture and Society*, 66, 17–27. <https://doi.org/10.26650/JECS2022-1116979>
- Philpot, R., Lindquist, M., Mullis, I. V. S., & Aldrich, C. E. A. (2021). TIMSS 2023 mathematics framework. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2023 assessment frameworks*. (pp. 5-19). Boston College, TIMSS & PIRLS International Study Center <https://timssandpirls.bc.edu/timss2023/frameworks/index.html>
- Polat, S., & Dede, Y. (2023). Trends in cognitive demands levels of mathematical tasks in Turkish middle school mathematics textbooks: Algebra learning domain. *International Journal for Mathematics Teaching and Learning*, 24(1), 40–61. <https://doi.org/10.4256/ijmtl.v24i1.476>
- R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reilly, D., Neumann, D. L., & Andrews, G. (2019). Investigating gender differences in mathematics and science: Results from the 2011 Trends in Mathematics and Science Survey. *Research in Science Education*, 49, 25–50. <https://doi.org/10.1007/s11165-017-9630-6>
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. <https://doi.org/10.1002/9780470316696>
- Şahin, A., Kurt, G., & Yerdelen-Damar, S. (2025). Factors predicting Turkish students' mathematics and science achievement in TIMSS 2019. *Electronic Journal for Research in Science & Mathematics Education*, 29(1), 42–59.

- Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (Eds.). (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics* (Vol. 1). Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-011-5786-5>
- Şensoy, Ç. P., & Kılıç, A. (2021). Ortaokul matematik öğretmenlerinin ders işleyiş süreçlerinin incelenmesi [Analysis of the middle school math teachers' teaching processes]. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 21(2), 434–452. <https://doi.org/10.17240/aibuefd.2021.21.62826-874046>
- Sezer, E., & Çakan, M. (2022). Role of teacher quality and working conditions in TIMSS 2019 mathematics achievement. *Journal of Theoretical Educational Sciences*, 15(2), 395–419. <https://doi.org/10.30831/akueg.971286>
- Suna, H. E., & Özer, M. (2021). Türkiye'de sosyoekonomik düzey ve okullar arası başarı farklarının akademik başarı ile ilişkisi [The achievement gap between schools and relationship between achievement and socioeconomic status in Turkey]. *Journal of Measurement and Evaluation in Education and Psychology*, 12(1), 54–70. <https://doi.org/10.21031/epod.860431>
- Suna, H. E., Tanberkan, H., Gür, B. S., Perc, M., & Özer, M. (2020). Socioeconomic status and school type as predictors of academic achievement. *Journal of Economy Culture and Society*, 61, 41–64. <https://doi.org/10.26650/JECS2020-0034>
- Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, 34(1), 89–101. <https://doi.org/10.1016/j.cedpsych.2008.09.002>
- Uzun, H., & Ağaç, G. (2023). Ortaokul matematik öğretmenlerinin beceri temelli sorulara ilişkin yaklaşımlarının incelenmesi [An investigation of secondary school mathematics teachers' approaches to skill-based question]. *İnönü University Journal of the Faculty of Education*, 24(1), 57–78. <https://doi.org/10.17679/inuefd.1133896>
- von Davier, M., Kennedy, A., Reynolds, K., Fishbein, B., Khorramdel, L., Aldrich, C., Bookbinder, A., Bezirhan, U., & Yin, L. (2024). *TIMSS 2023 international results in mathematics and science*. TIMSS & PIRLS International Study Center, Boston College. <https://doi.org/10.6017/lse.tpisc.timss.rs6460>
- Yalçın, S. (2022). Examining students' item response times in eTIMSS according to their proficiency levels, self-confidence, and item characteristics. *Journal of Measurement and Evaluation in Education and Psychology*, 13(1), 23–39. <https://doi.org/10.21031/epod.999545>
- Yıldırım, S. (2019). Predicting mathematics achievement: The role of socioeconomic status, parental involvement, and self-confidence. *Education and Science*, 44(198), 99–113. <https://doi.org/10.15390/EB.2019.7868>
- Yin, L., & Foy, P. (2021). TIMSS 2023 assessment design. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2023 assessment frameworks*. (pp. 71–85). Boston College, TIMSS & PIRLS International Study Center <https://timssandpirls.bc.edu/timss2023/frameworks/index.html>
- Zhu, Z. (2023). International comparative study of learning trajectories based on TIMSS 2019 G4 data on cognitive diagnostic models. *Frontiers in Psychology*, 14, 1241656. <https://doi.org/10.3389/fpsyg.2023.1241656>