University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 21/2026 (515)

Beyond Weeks Closed: Reassessing Pandemic Educational Disruption with TIMSS 2003–2023

Marzena Binkiewicz
Artur Pokropek

Warsaw 2026
ISSN 2957-0506

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

Beyond Weeks Closed: Reassessing Pandemic Educational Disruption with TIMSS 2003–2023

Marzena Binkiewicz[1], Artur Pokropek[2]*

[1] University of Warsaw, Faculty of Economic Sciences
[2] Polish Academy of Sciences, Institute of Philosophy and Sociology
[*] Corresponding author: mn.binkiewicz@uw.edu.pl

Abstract: Using six TIMSS cycles (2003–2023), UNESCO school-closure data, and World Bank GDP indicators, we estimate deviation-from-trend and continuous-treatment Difference-in-Differences models for mathematics achievement, liking learning mathematics, and confidence in mathematics. Mathematics achievement declined substantially relative to pre-pandemic trends, but closure duration did not consistently explain cross-national variation in achievement losses. Unexpectedly, longer closures were associated with slightly higher liking learning mathematics, particularly among Grade 4 students in higher-GDP countries, while confidence in mathematics remained largely unchanged. These findings suggest a decoupling of cognitive and affective outcomes during the pandemic and indicate that school-closure duration alone is an insufficient measure of educational harm. Comparative evaluations of pandemic education policies should therefore incorporate indicators of instructional continuity, remote-learning quality, and recovery efforts alongside formal school-closure measures.

Keywords: Mathematics achievement, non-cognitive factors, distance learning, school closures, TIMSS, math anxiety, self-efficacy

JEL codes: I21, I24, C21, C23, C83

1. Introduction

When school systems worldwide closed in 2020, the cumulative duration of full school closures rapidly became the dominant cross-national indicator of pandemic disruption — used to rank policy responses, project learning losses, and motivate recovery spending (Betthäuser et al., 2023; Patrinos et al., 2023; UNESCO, 2021). The assumption underlying this metric is that longer school closures are associated with greater educational disruption and therefore worse educational outcomes. A substantial literature has documented post-pandemic declines in student achievement consistent with that view (Engzell et al., 2021; Jakubowski et al., 2025; Kuhfeld et al., 2020), particularly in mathematics (Betthäuser et al., 2023; Gajderowicz et al., 2026; Di Pietro, 2023). Yet whether the data actually support the closure-duration assumption, especially beyond cognitive outcomes and outside the early-studied OECD sample, has rarely been tested directly. Considerably less is known, in particular, about whether pandemic disruptions reshaped students' broader relationship with mathematics, including their enjoyment of the subject and their confidence in learning it.

Substantial declines in academic achievement appear especially pronounced in lower-income settings (Dela Cruz et al., 2025; Jakubowski et al., 2025; Székely et al., 2024), where students often faced limited internet access, fewer educational resources outside school, and less supportive home-learning environments during remote instruction (Alexander et al., 2001; Lichand et al., 2021). At the same time, education systems differed markedly in their institutional capacity to respond to the pandemic (Kastorff & Heine, 2025; Reimers & Schleicher, 2020; Zancajo et al., 2022). While economically more developed systems generally adopted remote learning more rapidly, lower-income systems faced weaker digital infrastructure and greater pandemic vulnerability (Pejić Bach et al., 2023).

Educational systems produce not only cognitive skills but also motivational and affective orientations toward learning. In mathematics education, factors such as confidence, self-efficacy and intrinsic motivation are strongly associated with academic achievement (Bandura, 1997; Lee, 2009; Marsh et al., 2005). These associations are comparable in magnitude to major socioeconomic predictors (Lee & Stankov, 2018), while remaining potentially more responsive to instructional practices and school-based interventions (Lee & Shute, 2010).

If the pandemic disrupted classroom routines, relationships with teachers and peers, and students' mastery experiences, theory would predict a deterioration in mathematics-related orientations alongside declines in achievement (Bandura, 1997; Kalogeropoulos et al., 2021).

Empirical evidence, however, remains mixed. Some studies report increases in mathematics anxiety and declines in motivation under remote instruction (Li et al., 2023; Lichand et al., 2021; Quintero Pena et al., 2025), whereas others find heterogeneous or even neutral patterns depending on instructional conditions and student characteristics (Doz & Doz, 2022). Still other studies highlight circumstances under which remote and self-paced learning formats supported, rather than undermined, student engagement (Edwards & Rule, 2013; Hwang et al., 2021). Much of this evidence, however, relies on small or context-specific samples, limiting broader comparative inference. Whether school closures systematically weakened — or in some contexts strengthened — students' relationship with mathematics, a subject central to post-pandemic recovery policy, therefore remains an open question.

This study addresses these gaps using six cycles of the Trends in International Mathematics and Science Study (TIMSS), spanning 2003–2023. TIMSS enables the evaluation of post-pandemic deviations from long-term pre-pandemic trajectories using harmonized measures repeatedly collected across countries and two grade levels. The widespread use of simple cross-national indicators reflects a broader trend in comparative education policy toward standardized and easily comparable measures of educational performance and disruption (Addey et al., 2017). We also combine TIMSS data with UNESCO indicators of school-closure duration (UNESCO, 2021) and World Bank measures of GDP per capita to examine whether prolonged school closures were systematically associated with post-pandemic changes in mathematics achievement and mathematics-related non-cognitive outcomes across countries with different levels of economic development.

This study contributes to comparative education policy research in three ways. First, it evaluates whether closure duration is a robust predictor of post-pandemic outcomes accounting for long-term pre-pandemic trajectories. Second, it extends the literature on pandemic schooling beyond achievement by examining mathematics-related non-cognitive outcomes in a comparative perspective. Third, it examines heterogeneity by country economic context, addressing broader questions about how educational systems with different levels of economic resources respond to large-scale disruptions (Pejić Bach et al., 2023).

Based on the existing literature, we formulate three expectations regarding our empirical analyses. First, drawing on evidence that educational losses during COVID-19 were substantially larger in lower-income countries due to limited access to digital infrastructure and inadequate learning environments (Lichand et al., 2021), we expect prolonged school closures to be associated with larger declines in mathematics achievement in less economically developed contexts. Second, the effects of remote learning on students' attitudes toward mathematics may have been heterogeneous. While reduced instructional interaction may have weakened engagement for some students, greater flexibility and self-paced learning environments may have benefited others (Edwards & Rule, 2013). Third, following Bandura's (1997) social cognitive theory, we expect confidence in mathematics also to be sensitive to prolonged remote instruction because distance learning likely reduced opportunities for direct teacher support, classroom participation, and mastery experiences.

2. Data and methods

2.1. Data

This study uses data from six cycles of the Trends in International Mathematics and Science Study (TIMSS), spanning 2003–2023. TIMSS is a repeated cross-national assessment of mathematics achievement conducted among Grade 4 and Grade 8 students using harmonized sampling procedures and internationally standardized assessment frameworks (von Davier et al., 2024). Countries were included if they participated in both TIMSS 2019 and TIMSS 2023 and in at least two earlier cycles, allowing the estimation of pre-pandemic trajectories and placebo tests. The Grade 4 sample includes approximately 1 million student observations from 37 countries, while the Grade 8 sample includes approximately 990,000 observations from 29 countries. Country coverage is reported in Supplementary Appendix A.

TIMSS uses a two-stage stratified sampling design in which schools are sampled first and classes are sampled within schools. In addition to achievement measures, TIMSS collects extensive contextual information from students, teachers, and principals regarding instructional environments and home educational resources. Because mathematics-related non-cognitive measures are available only from TIMSS 2011 onward, analyses of these outcomes rely on shorter pre-treatment periods and exclude a small number of countries with insufficient observations for placebo estimation (Grade 4: CYP and LVA; Grade 8: CYP, KWT, and ROU).

To capture pandemic-related institutional disruption, TIMSS data are linked with UNESCO indicators of full school-closure duration and World Bank measures of GDP per capita. School-closure data cover the period from February 2020 through March 2022 and record the cumulative duration of government-mandated full school closures. GDP per capita is measured in 2019 U.S. dollars and serves as an indicator of countries' pre-pandemic level of economic development.

2.2. Measures

Mathematics achievement. Mathematics achievement is measured using TIMSS plausible values, which represent multiple imputations of students' latent achievement and account for measurement uncertainty associated with large-scale assessments. The TIMSS mathematics framework defines achievement in relation to content domains and cross-cutting cognitive domains, with content domains covering number, measurement, and geometry, and data at the fourth grade and number, algebra, geometry and measurement, and data and probability at the eighth grade. TIMSS achievement scores are scaled using item response theory procedures and standardized to an international mean of 500 and standard deviation of 100.

Mathematics-related non-cognitive outcomes. Two TIMSS scales are used to capture students' relationship with mathematics. The first, *Students Like Learning Mathematics*, measures students' enjoyment of and engagement with mathematics learning. The second, *Students Confident in Mathematics*, captures students' perceived competence and mathematics self-confidence. Both scales are standardized to an international mean of 10 and a standard deviation of 2.

These outcomes are theoretically relevant because motivational orientations toward mathematics are strongly associated with academic achievement and educational participation (Eccles & Wigfield, 2002; Marsh & Martin, 2011). Unlike structural socioeconomic conditions, they are also considered comparatively malleable through instructional practices and learning environments (Lee & Shute, 2010).

School-closure duration. The main explanatory variable is the cumulative duration of full school closures, measured in weeks using UNESCO data. Globally, schools remained fully closed for an average of approximately five months (20 weeks), whereas closure durations in the analysed TIMSS samples were shorter, averaging between 16 and 18 weeks (Table 1).

Table 1. Descriptive Statistics for Full School Closure Duration and GDP per Capita Across Analytical Samples

Panel A. Full School Closure Duration (weeks)

Sample	P25	Median	P75	Mean	SD	N countries
UNESCO	10	16	27	20	15	210
TIMSS Grade 4	8	13	20	16	13	37
TIMSS Grade 8	8	15	22	18	15	29

Panel B. GDP per Capita in 2019 (USD)

Sample	P25	Median	P75	Mean	SD	N countries
UNESCO	2,190	6,582	20,584	17,735	27,104	198
TIMSS Grade 4	17,013	29,703	48,358	33,149	21,882	37
TIMSS Grade 8	10,920	29,567	44,251	30,087	21,741	29

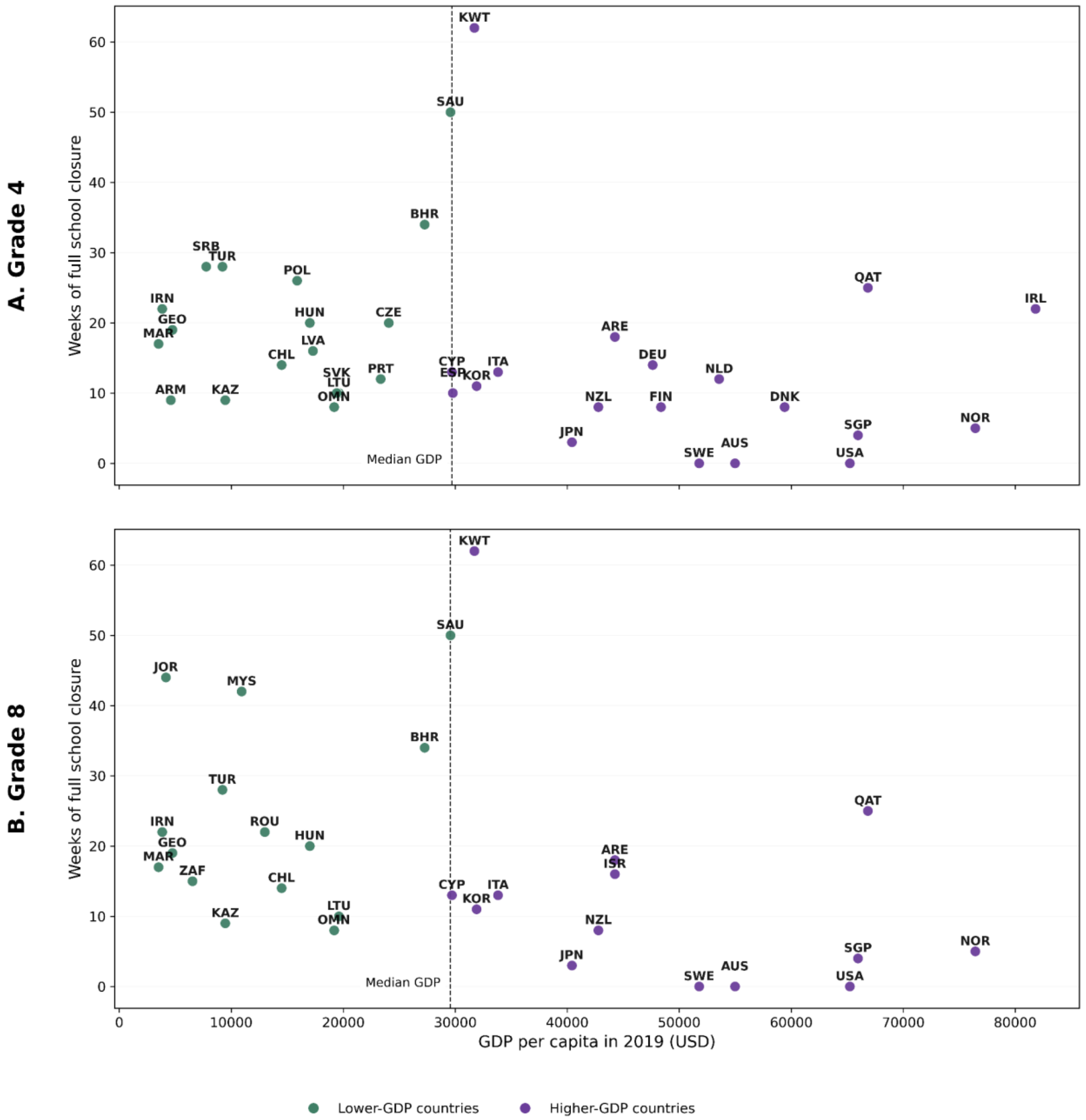
Note: GDP values are reported in current U.S. dollars. P25 = 25th percentile; P75 = 75th percentile; SD = standard deviation.

Country development status. To examine cross-national heterogeneity, countries are classified into lower- and higher-GDP groups based on median GDP per capita within each analytical sample in 2019 (Figure 1).

To assess the sensitivity of the results to the choice of threshold, supplementary analyses additionally used alternative GDP classifications based on the 40th and 60th percentiles of the 2019 GDP-per-capita distribution. The overall pattern of findings was substantively unchanged (see Suplementary Appendix E).

Kuwait represents an extreme case of prolonged closure duration and therefore constitutes a potentially influential observation. The country is retained in the main analyses because it reflects a substantive policy response rather than a measurement anomaly. Additional robustness checks, excluding Kuwait and capping treatment intensity, are reported in Supplementary Appendix F.

Figure 1. GDP per Capita and Duration of Full School Closures in TIMSS Samples.



Individual-level controls. All models include student gender, age, and home educational resources. The latter is represented either by number of books in models concerning mathematics or by a TIMSS composite measure based on, among others, the number of books at home, own room and internet connection (von Davier et al., 2024).

Descriptive statistics are broadly comparable across grade levels; however, the home educational resources variable contains substantial missing data, particularly in the Grade 4 sample (Table 2). Missing values are addressed using multiple imputation by chained equations

(Bouhlila & Sellaouti, 2013; van Buuren et al., 1999), with the imputation model additionally incorporating the five mathematics plausible values.

Table 2. Descriptive Statistics for Individual-Level Variables

Panel A. Grade 4

Variable	N	Mean	SD	Min	Max
Gender (1=female)	1001074	1.51	0.50	1	2
Age	999280	10.22	0.60	6	15
Home resources for learning	578022	10.16	1.83	3.61	15.14

Panel B. Grade 8

Variable	N	Mean	SD	Min	Max
Gender (1=female)	982134	1.51	0.50	1	2
Age	980068	14.29	0.80	9	20.33
Home resources for learning	745961	10.18	1.82	4.23	14.02

2.3. Methods

The empirical strategy exploits cross-national variation in the duration of full school closures during the COVID-19 pandemic to examine whether countries with longer periods of closure experienced systematically different post-pandemic trajectories in mathematics achievement and mathematics-related non-cognitive outcomes.

The first specification is a deviation-from-trend model. It serves two purposes. First, it estimates whether outcomes in 2023 deviated from country-specific pre-pandemic trends. Second, it examines whether these deviations varied systematically with the duration of school closures. This specification compares outcomes observed in 2023 with values projected from country-specific pre-pandemic trajectories (Jakubowski et al, 2025; Kennedy & Strietholt, 2023):

$$Y_{ic} = \sum a_c + \sum \beta_c * time + \rho POST_{2023} + \sigma POST_{2023} * weeks_c + \gamma X_{ic} + \varepsilon_{ic} \quad ,$$

where α denotes country fixed effects, β captures country-specific pre-pandemic trends, and $POST_{2023}$ is a binary indicator for the TIMSS 2023 cycle. In models that incorporate continuous treatment intensity, the post-period indicator is interacted with $weeks_c$, defined as the number of weeks during which schools were fully closed because of COVID-19

The second specification is a continuous-treatment Difference-in-Differences (DiD) model in which treatment intensity is measured by the duration of school closures (Angrist and Pischke, 2008):

$$Y_{ict} = a_c + \lambda_t + \delta\, weeks_c * POST_{2023} + \gamma X_{ict} + \varepsilon_{ict},$$

where $weeks_c$ measures closure duration in weeks and δ captures whether post-pandemic changes vary systematically with treatment intensity.

Following Callaway et al. (2024), continuous-treatment DiD models involve two distinct identifying assumptions. The first concerns average treatment effects across treatment levels and relies on a generalized parallel trends assumption. The second concerns causal responses to additional treatment intensity and requires that marginal treatment effects are homogeneous across units. We assess the plausibility of the generalized parallel trends assumption by examining whether countries exposed to different levels of treatment intensity exhibited systematically different pre-pandemic trends (Angrist and Pischke, 2008).

Nevertheless, there are substantial differences between countries with shorter and longer school closures, particularly with respect to GDP per capita (Figure 1). Such heterogeneity raises concerns about the assumption that an additional week of school closure has the same marginal effect across countries. Consequently, causal interpretations of treatment intensity should be made with caution (Callaway et al., 2024). As an additional diagnostic, all main specifications are re-estimated using a placebo treatment period between 2015 and 2019, excluding observations from 2023. Estimates that fail either the pre-trend diagnostic or the placebo test are interpreted as insufficient to support causal claims regarding the effects of school-closure duration.

Following standard TIMSS procedures, all analyses were estimated separately for each plausible value and combined using Rubin's multiple-imputation rules to account for both sampling variance and imputation variance (Little & Rubin, 1987; von Davier et al., 2024). Point estimates are computed using senate weights to ensure equal weighting of participating countries rather than weighting estimates by national student population size. This approach aligns with the study's comparative focus on cross-national variation in educational disruption rather than population-average effects across students. Standard errors are clustered at the country level to account for serial correlation within countries across assessment cycles.

3. Results

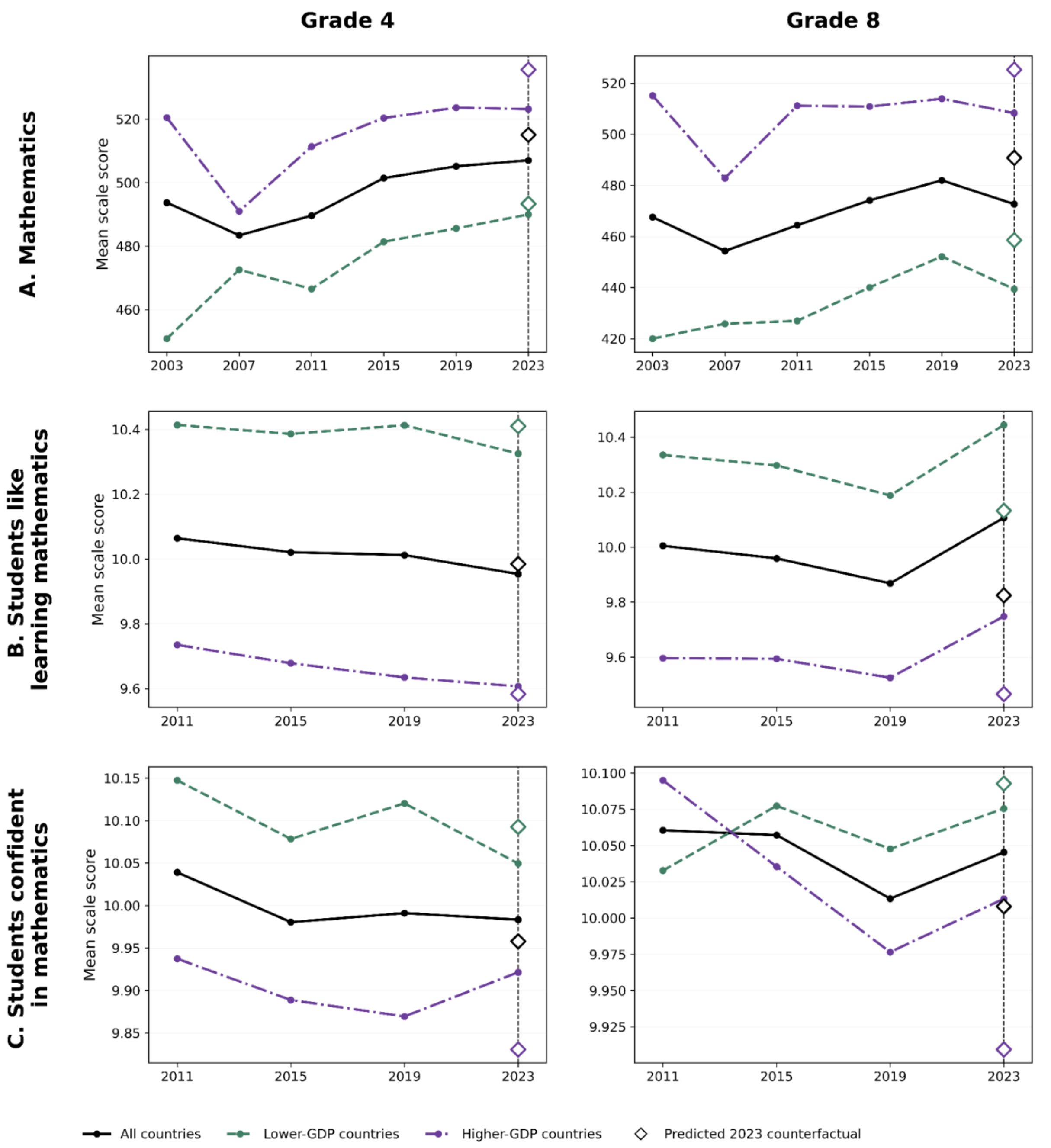
3.1. Post-pandemic deviations from trend

We first compare post-pandemic outcomes with pre-pandemic trajectories and then assess whether cross-national variation in school-closure duration helps explain these patterns. Figure 2 presents observed trends across TIMSS cycles for Grades 4 and 8 alongside predicted 2023 counterfactual values derived from country-specific pre-pandemic linear trends. Comparing observed and predicted outcomes provides a descriptive assessment of the extent to which 2023 outcomes deviated from historical trajectories. Counterfactual predictions are estimated using data from cycles prior to 2023, with standard errors clustered at the country level. Results are presented separately for all participating countries, countries with lower GDP per capita, and countries with higher GDP per capita. The visual patterns are consistent with the regression estimates reported in Supplementary Appendix B.

Across both grades, mathematics achievement declined relative to pre-pandemic trajectories, with estimated deviations of approximately −0.09 standard deviations (SD) in Grade 4 and −0.18 SD in Grade 8. These declines are consistent with prior evidence of substantial pandemic-related learning losses, particularly in mathematics (Bertoletti et al., 2023; Borgonovi & Ferrara, 2023; Wisenöcker et al., 2025). However, the magnitude of the decline varies across grades and country-income groups.

In Grade 4, deviation-from-trend estimates indicate significant losses primarily in higher-GDP countries (-0.14 SD), whereas losses in lower-GDP countries are negative but not statistically significant. One possible explanation is that higher-GDP countries experienced stronger pre-pandemic achievement growth, implying that a stabilization of outcomes after the pandemic would generate larger departures from projected trajectories. In Grade 8, achievement losses are substantial in both lower-GDP countries (-0.18 SD) and higher-GDP ones (-0.17 SD). The difference between the two groups is not statistically significant.

Figure 2. Observed 2023 Outcomes Relative to Predicted Pre-Pandemic Trends, by Grade and Country GDP Group.



The pattern for mathematics-related non-cognitive outcomes differs from that observed for achievement. Students' liking learning mathematics generally remained stable or improved relative to pre-pandemic trends, particularly in Grade 8, where positive deviations are statistically significant for all (+0.16 SD), lower-GDP (+0.18 SD) and higher-GDP (+0.14 SD) countries. Changes in students' confidence in mathematics are smaller and less consistent. Although most estimates are positive, they do not reach statistical significance at p=0.05.

Taken together, these findings suggest a partial decoupling between cognitive and non-cognitive outcomes: achievement declined relative to pre-prandemic trends, whereas mathematics-related attitudes and confidence either remained stable, or improved.

3.2. School closure duration as a policy indicator

The central policy question is whether the duration of full school closures explains cross-national differences in post-pandemic educational and non-educational outcomes. Overall, the evidence suggests that closure duration is not a consistent predictor of cross-national variation in achievement losses (Table 3). First, effects for both specifications either are small and not statistically significant or fail placebo tests. Second, the generalized parallel trends assumption required for causal interpretation of continuous-treatment Difference-in-Differences estimates is not met (see Supplementary Appendix D).

Table 3. School-Closure Duration and Post-Pandemic Mathematics Outcomes: Main and Placebo Estimates

Grade	Model	All countries	Poorer countries	Richer countries
Grade 4	DiD	0.624*** (0.185)	0.595 (0.414)	0.692*** (0.240)
	DiD Placebo	0.608** (0.243)	0.335 (0.462)	0.894*** (0.226)
	Trend	-0.014 (0.276)	0.145 (0.576)	-0.302 (0.268)
	Trend Placebo	0.292 (0.244)	0.276 (0.562)	0.184 (0.350)
Grade 8	DiD	-0.106 (0.357)	-0.436 (0.720)	0.336 (0.279)
	DiD Placebo	0.268 (0.203)	0.219 (0.405)	0.426 (0.261)
	Trend	-0.264 (0.210)	-0.200 (0.361)	-0.413 (0.302)
	Trend Placebo	0.356 (0.292)	0.926 (0.598)	-0.444 (0.418)

*Notes: * p<0.10, ** p<0.05, *** p<0.01.*

Only for Grade 4, the continuous-treatment DiD models suggest a positive association between school-closure duration and mathematics achievement. However, these estimates fail placebo tests and generalized PTA diagnostics. Taken together, findings provide little evidence that the number of weeks schools were fully closed can, by itself, account for cross-national variation in post-pandemic mathematics achievement.

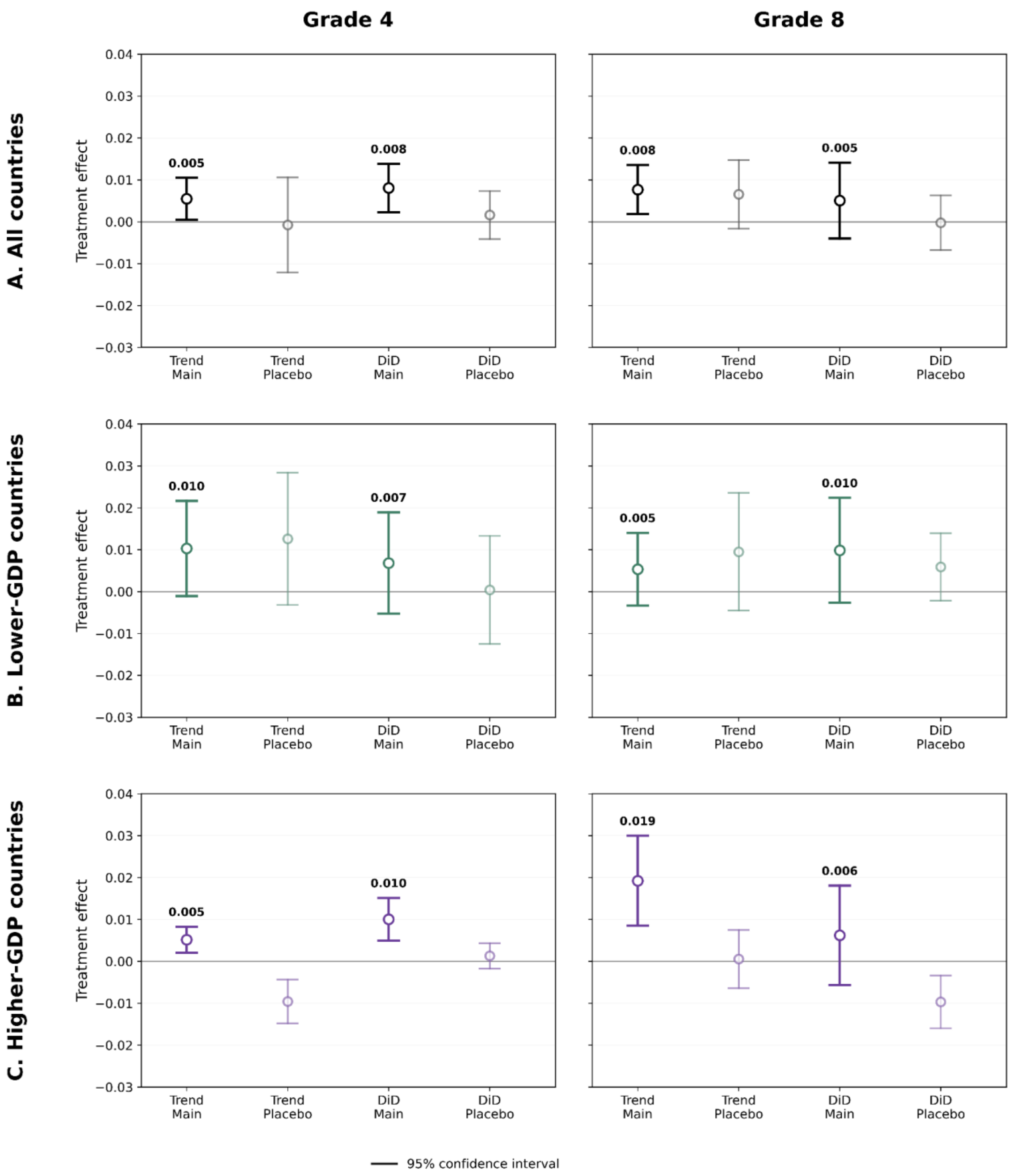
A much more consistent closure-duration pattern concerns students' liking learning mathematics, for which estimates from all specifications, together with 95% confidence intervals, are presented in Figure 3. In Grade 4, both the continuous-treatment DiD and deviation-from-trend models indicate a small positive association between weeks of school closure and liking learning mathematics, corresponding to approximately 0.003–0.004 SD per additional week. The association is most pronounced among higher-GDP countries (+0.005 SD). In Grade 8, the DiD estimates are generally positive but statistically insignificant, while the deviation-from-trend estimates are positive and significant for the full sample (+0.004 SD) and higher-GDP countries (+0.01 SD).

Overall, the results suggest that countries experiencing longer periods of school closure tended to exhibit more favorable post-pandemic trajectories in students' reported liking learning mathematics, with somewhat stronger evidence among younger students and in higher-GDP contexts.

Evidence regarding students' confidence in mathematics is more limited (see Supplementary Appendix C). In Grade 4, the continuous-treatment DiD estimates indicate a small negative association between school-closure duration and confidence (-0.006 SD) that reaches statistical significance only at the 10% level. In contrast, the deviation-from-trend specification yields positive and statistically significant estimates only for higher-GDP countries (+0.005 SD). In Grade 8, DiD estimates are close to zero, whereas the deviation-from-trend model similarly produces positive and statistically significant estimates for higher-GDP countries (+0.017 SD). Although these results point to modest improvements in confidence relative to pre-pandemic trends among higher-GDP countries, the lack of consistency across specifications warrants cautious interpretation.

Robustness checks using alternative GDP thresholds (see Supplementary Appendix E) support the general interpretation: the achievement and confidence results remain weak and inconsistent, while the estimates for liking mathematics are more stable, especially in higher-GDP contexts. Additional robustness checks excluding Kuwait and capping treatment intensity at 50 weeks yield substantively similar conclusions (see Supplementary Appendix F).

Figure 3. Estimated Effects of Pandemic-Related School Closures on Students' Liking Learning Mathematics



4. Conclusions and Discussion

This study examined whether COVID-19 school closures were associated with changes in mathematics achievement and two mathematics-related non-cognitive outcomes: students' liking learning mathematics and confidence in learning mathematics. Using six TIMSS cycles and two complementary modeling strategies, we identify three main findings.

First, mathematics achievement declined substantially relative to pre-pandemic trajectories, but closure duration does not robustly explain these declines. This is an important policy finding. While recent cross-national TIMSS evidence has reported larger learning losses in countries experiencing longer school closures (Gajderowicz et al., 2026), our results suggest that these relationships become substantially weaker when analyses are restricted to countries with longer pre-treatment histories and are subjected to alternative specifications, placebo tests, and other validity checks. Much of the public debate treated the number of weeks of school closure as a direct measure of educational harm. Our results suggest that this indicator is insufficient in the cross-national TIMSS sample. Cross-national differences in institutional capacity, digital preparedness, instructional quality, and recovery policies likely mattered at least as much as the duration of full closure itself.

Second, students' liking learning mathematics did not decline alongside achievement. Instead, it generally remained stable or improved relative to pre-pandemic trends, particularly in Grade 8, where positive deviations were observed among all countries as well as among lower- and higher-GDP groups. Moreover, longer school closures were associated with slightly higher levels of reported liking for learning mathematics, especially in Grade 4 and among higher-GDP countries. These effects suggest that pandemic schooling may have affected cognitive and affective outcomes differently. One possible explanation is that some students experienced remote or hybrid learning as less pressured and more self-paced, particularly in educational systems with stronger digital infrastructure and more supportive home-learning environments (Borba, 2021; Hwang et al., 2021). However, because TIMSS does not directly measure the quality of remote instruction, this mechanism should be treated as suggestive rather than conclusive.

Third, confidence in mathematics shows less robust evidence of change. A mild positive association with additional weeks of school closure is observed for both higher-GDP countries in both grades, but this pattern is supported by only one modelling approach. Confidence in learning mathematics may depend more directly on mastery experiences, teacher feedback, classroom participation, and social reinforcement, all of which may have been harder to sustain during remote instruction (Bandura, 1997; Marsh & Martin, 2011). These findings suggest that enjoyment of mathematics and confidence in mathematics should not be treated as interchangeable dimensions of student engagement.

The results also point to developmental differences. Evidence for associations between school-closure duration and affective outcomes was stronger among younger students, whereas results for Grade 8 were generally weaker and less robust across specifications. At the same time, declines in mathematics achievement after the pandemic were substantially larger among older students. This pattern is consistent with research showing that mathematics interest, enjoyment, and perceived competence tend to decline across adolescence and are closely linked to achievement trajectories (Jacobs et al., 2002; Eccles & Wigfield, 2020). Younger students' attitudes toward mathematics may therefore be more malleable and more responsive to changes in learning environments.

Several limitations should guide interpretation. First, the analysis relies on country-level measures of full school-closure duration and therefore cannot capture within-country variation in reopening policies, hybrid instruction, remote-learning quality, teacher practices, or household access to digital resources (Alfano, 2022; Zancajo et al., 2022). Second, causal interpretations of closure duration are constrained by the assumptions underlying continuous-treatment models, particularly the assumption of homogeneous marginal treatment effects across units (Callaway et al., 2024). Finally, TIMSS 2023 captures medium-term post-pandemic outcomes rather than immediate effects of school closures.

The policy implication is not that school closures were harmless. Rather, the results caution against treating closure duration as a stand-alone measure of educational disruption. Future large-scale disruptions to schooling, whether caused by pandemics, environmental crises, or political instability, remain plausible policy challenges (Peiris, 2025). Evaluations of crisis-related schooling policies should therefore combine closure indicators with measures of instructional continuity, remote-learning quality, access to digital resources, teacher support, compensatory policies, and recovery interventions. From a policy perspective, the quality of learning support available during school closures may matter more than the duration of closures itself (Angrist et al., 2022).

References

Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B., Verger, A., 2017. The rise of international large-scale assessments and rationales for participation. Compare: A Journal of Comparative and International Education, 47(3), 434–452. https://doi.org/10.1080/03057925.2017.1301399

Alexander, K.L., Entwisle, D.R., Olson, L.S., 2001. Schools, achievement, and inequality: A seasonal perspective. Educational Evaluation and Policy Analysis, 23(2), 171–191. https://doi.org/10.3102/01623737023002171

Alfano, V., 2022. The effects of school closures on COVID-19: A cross-country panel analysis. Applied Health Economics and Health Policy, 20(2), 223–233. https://doi.org/10.1007/s40258-021-00702-z

Angrist, J.D., Pischke, J.-S., 2008. Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

Angrist, N., Bergman, P., Matsheng, M., 2022. Experimental evidence on learning using low-tech when school is out. Nature human behaviour, 6(7), 941-950.

Bandura, A., 1997. Self-efficacy: The exercise of control. W. H. Freeman.

Bertoletti, A., Cannistrà, M., Soncin, M., Agasisti, T., 2023. The heterogeneity of COVID-19 learning loss across Italian primary and middle schools. Economics of Education Review, 95, Article 102435. https://doi.org/10.1016/j.econedurev.2023.102435

Betthäuser, B.A., Bach-Mortensen, A.M., Engzell, P., 2023. A systematic review and meta-analysis of evidence on learning during the COVID-19 pandemic. Nature Human Behaviour, 7(3), 375–385. https://doi.org/10.1038/s41562-022-01506-4

Borba, M.C., 2021. The future of mathematics education since COVID-19. Educational Studies in Mathematics, 108, 385–400. https://doi.org/10.1007/s10649-021-10043-2

Borgonovi, F., Ferrara, A., 2023. COVID-19 and inequalities in educational achievement in Italy. Research in Social Stratification and Mobility, 83, Article 100760. https://doi.org/10.1016/j.rssm.2023.100760

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(4). https://doi.org/10.1186/2196-0739-1-4

Callaway, B., Goodman-Bacon, A., Sant'Anna, P.H.C., 2024. Difference-in-differences with a continuous treatment (NBER Working Paper No. 32117). National Bureau of Economic Research. https://doi.org/10.3386/w32117

Dela Cruz, N.A., Adona, A.J., Molato-Gayares, R., Park, A., 2025. Learning loss and recovery from the COVID-19 pandemic: A systematic review of evidence. International Journal of Educational Development, 115, 103271. https://doi.org/10.1016/j.ijedudev.2025.103271

Di Pietro, G., 2023. The impact of COVID-19 on student achievement: Evidence from a recent meta-analysis. Educational Research Review, 39, Article 100530. https://doi.org/10.1016/j.edurev.2023.100530

Doz, D., Doz, E., 2022. The impact of COVID-19 distance learning on students' math anxiety: An exploratory study. International Journal of Education in Mathematics, Science and Technology, 11(1), 1–16. https://doi.org/10.46328/ijemst.2219

Eccles, J.S., Wigfield, A., 2002. Motivational beliefs, values, and goals. Annual Review of Psychology, 53, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

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

Edwards, C., Rule, A.C., 2013. Attitudes of middle school students: Learning online compared to face to face. Journal of Computers in Mathematics and Science Teaching, 32(1), 49–66.

Engzell, P., Frey, A., Verhagen, M.D., 2021. Learning loss due to school closures during the COVID-19 pandemic. Proceedings of the National Academy of Sciences, 118(17), e2022376118. https://doi.org/10.1073/pnas.2022376118

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

Hwang, G.-J., Wang, S.-Y., Lai, C.-L., 2021. Effects of a social regulation-based online learning framework on students' learning achievements and behaviors in mathematics. Computers & Education, 160, Article 104031. https://doi.org/10.1016/j.compedu.2020.104031

Jacobs, J.E., Lanza, S., Osgood, D.W., Eccles, J.S., Wigfield, A., 2002. Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. Child Development, 73(2), 509–527. https://doi.org/10.1111/1467-8624.00421

Jakubowski, M., Gajderowicz, T., & Patrinos, H. A. (2025). COVID-19, school closures, and student learning outcomes: New global evidence from PISA. npj Science of Learning, 10, Article 1. https://doi.org/10.1038/s41539-025-00297-3

Kalogeropoulos, P., Roche, A., Russo, J., Vats, S., Russo, T., 2021. Learning mathematics from home during COVID-19: Insights from two inquiry-focussed primary schools. Eurasia Journal of Mathematics, Science and Technology Education, 17(5), em1957. https://doi.org/10.29333/ejmste/10830

Kastorff, T., Heine, J., 2025. Socioeconomic status and scientific literacy: Expanding educational inequalities during the COVID-19 pandemic—Insights from PISA 2018 and 2022. Large-scale Assessments in Education, 13, 36. https://doi.org/10.1186/s40536-025-00273-8

Kennedy, A.I., Strietholt, R., 2023. School closure policies and student reading achievement: Evidence across countries. Educational Assessment, Evaluation and Accountability, 35(4), 475–501. https://doi.org/10.1007/s11092-023-09415-4

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., Liu, J., 2020. Projecting the potential impact of COVID-19 school closures on academic achievement. Educational Researcher, 49(8), 549–565. https://doi.org/10.3102/0013189X20965918

Lee, J., 2009. Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. Learning and Individual Differences, 19(3), 355–365. https://doi.org/10.1016/j.lindif.2008.10.009

Lee, J., Borgonovi, F., 2022. Relationships between family socioeconomic status and mathematics achievement in OECD and non-OECD countries. Comparative Education Review, 66(2), 199–227. https://doi.org/10.1086/719281

Lee, J., Shute, V.J., 2010. Personal and social-contextual factors in K–12 academic performance: An integrative perspective on student learning. Educational Psychologist, 45(3), 185–202. https://doi.org/10.1080/00461520.2010.493471

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

Li, D., Liew, J., Raymond, D., Hammond, T., 2023. Math anxiety and math motivation in online learning during stress: The role of fearful and avoidance temperament and implications for STEM education. PLOS ONE, 18(12), Article e0292844. https://doi.org/10.1371/journal.pone.0292844

Lichand, G., Dória, C.A., Leal Neto, O., Cossi, J., 2021. The impacts of remote learning in secondary education: Evidence from Brazil during the pandemic. Inter-American Development Bank. https://doi.org/10.18235/0003344

Little, R.J.A., Rubin, D.B., 1987. Statistical analysis with missing data.

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

Marsh, H.W., Trautwein, U., Lüdtke, O., Köller, O., Baumert, J., 2005. Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. Child Development, 76(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Pejić Bach, M., Jaković, B., Jajić, I., Meško, M., 2023. Investigating the Impact of COVID-19 on E-Learning: Country Development and COVID-19 Response. Mathematics, 11(6), 1520. https://doi.org/10.3390/math11061520

Patrinos, H., Vegas, E., Carter-Rau, R., 2023. An analysis of COVID-19 student learning loss. In Oxford research encyclopedia of economics and finance. Oxford University Press. https://doi.org/10.1093/acrefore/9780190625979.013.893

Peiris, M., 2025. Pandemic preparedness: Learning from COVID-19. Sri Lankan Journal of Infectious Diseases, 15(1), 1–3. https://doi.org/10.4038/sljid.v15i1.8817

Quintero Pena, J.W., Volante, L., De Witte, K., 2025. The impact of the pandemic and school closures on non-cognitive characteristics: Evidence from PISA. International Journal of Educational Research Open, 9, 100489. https://doi.org/10.1016/j.ijedro.2025.100489

Reimers, F.M., Schleicher, A., 2020. A framework to guide an education response to the COVID-19 pandemic of 2020. OECD Publishing. https://doi.org/10.1787/6ae21003-en

Sirin, S.R., 2005. Socioeconomic status and academic achievement: A meta-analytic review of research. Review of Educational Research, 75(3), 417–453. https://doi.org/10.3102/00346543075003417

Székely, M., Flores-Seceña, I., Hevia, F., Calderón, D., 2024. Measuring learning losses from delayed return to school: Evidence from Mexico. International Journal of Educational Development, 106, 102998. https://doi.org/10.1016/j.ijedudev.2024.102998

UNESCO, 2021, March 1. UNESCO figures show two-thirds of an academic year lost on average worldwide due to COVID-19 school closures. https://www.unesco.org/en/articles/unesco-figures-show-two-thirds-academic-year-lost-average-worldwide-due-covid-19-school-closures

UNESCO Institute for Statistics. (n.d.). Global monitoring of school closures caused by COVID-19. Retrieved February 23, 2025, from https://covid19.uis.unesco.org/global-monitoring-school-closures-covid19/

van Buuren, S., Boshuizen, H.C., Knook, D.L., 1999. Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in Medicine, 18(6), 681–694. https://doi.org/10.1002/(SICI)1097-0258(19990330)18:6

von Davier, M., Fishbein, B., Kennedy, A.M. (Eds.), 2024. TIMSS 2023 technical report: Methods and procedures. IEA / TIMSS & PIRLS International Study Center.

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. Boston College, TIMSS & PIRLS International Study Center.

Wisenöcker, A.S., Helm, C., Große, C.S., Hübner, N., Zitzmann, S., 2025. A meta-analysis of students' academic learning losses over the course of the COVID-19 pandemic. Learning and Instruction, 98, Article 102111. https://doi.org/10.1016/j.learninstruc.2025.102111

Zancajo, A., Verger, A., Bolea, C., 2022. Digitalization and beyond: The effects of COVID-19 on post-pandemic educational policy and delivery in Europe. Policy and Society, 41(1), 111–128. https://doi.org/10.1093/polsoc/puab016

Online Supplementary Materials

Supplementary Appendix A. Country participation across TIMSS cycles, student sample sizes, GDP per capita, and duration of school closures.

Country	Grade 4 cycles	Grade 8 cycles	Grade 4 N	Grade 8 N	GDP per capita in 2019	Weeks fully closed
ARE	5–8	5–8	96573	89264	44251	18
ARM	3–8		30537		4597	9
AUS	3–8	3–8	31992	44256	54973	0
BHR	5–8	3–8	19426	28880	27260	34
CHL	5–8	3, 5–8	19052	27180	14496	14
CYP	3, 6-8	3-4, 6–8	17344	16122	29703	13
CZE	4–8		25529		24063	20
DEU	4–8		21022		47624	14
DNK	4–8		19609		59404	8
ESP	5–8		31718		29787	10
FIN	5–8		20186		48358	8
GEO	4–8	4–8	21024	22277	4741	19
HUN	3–8	3–8	27419	29230	17013	20
IRL	5–8		18236		81810	22
IRN	3–8	3–8	29916	33288	3831	22
ISR		3–8		28615	44251	16
ITA	3–8	3–8	25519	27388	33813	13
JOR		3–8		39437	4170	44
JPN	3–8	3–8	25887	26678	40416	3
KAZ	4–8	5–8	24373	20092	9457	9
KOR	5–8	3–8	17251	29835	31902	11

Country	Grade 4 cycles	Grade 8 cycles	Grade 4 N	Grade 8 N	GDP per capita in 2019	Weeks fully closed
KWT	4–8	4, 6-8	20461	17500	31708	62
LTU	3–8	3–8	25880	28983	19609	10
LVA	3-4, 7–8		16375		17295	16
MAR	3–8	3–8	36636	45227	3508	17
MYS		3–8		42644	10920	42
NLD	3–8		20940		53555	12
NOR	3–8	3–8	25152	30236	76431	5
NZL	3–8	3, 5–8	31108	27260	42779	8
OMN	5–8	4–8	34398	38154	19180	8
POL	5–8		19322		15875	26
PRT	5–8		18249		23343	12
QAT	4–8	4–8	27675	26681	66841	25
ROU		3–5, 7-8		22041	12992	22
SAU	5–8	3–8	19820	28203	29567	50
SGP	3–8	3–8	37110	34156	65952	4
SRB	5–8		17144		7756	28
SVK	4–8		25387		19406	10
SWE	4–8	3–8	22585	29794	51773	0
TUR	5–8	4–8	22504	28326	9215	28
USA	3–8	3–8	58189	55243	65228	0
ZAF		3, 5–8		66204	6534	15

Note. TIMSS cycles refer to the official TIMSS assessment numbering: Cycle 3 = 2003, Cycle 4 = 2007, Cycle 5 = 2011, Cycle 6 = 2015, Cycle 7 = 2019, and Cycle 8 = 2023. "Grade 4 N" and "Grade 8 N" report the total number of student observations across all TIMSS cycles.

Supplementary Appendix B. Estimates of post-treatment effects.

Grade	Model	All countries	Poorer countries	Richer countries	Poorer – richer
		1. Mathematics			
Grade 4	Main	-8.899*** (3.257)	-4.185 (4.888)	-13.001*** (4.225)	0.172
	Placebo	-5.211 (3.344)	-0.695 (4.542)	-8.723* (4.646)	0.217
Grade 8	Main	-18.115*** (3.793)	-19.235*** (5.696)	-16.920*** (5.127)	0.763
	Placebo	0.930 (4.097)	6.777 (6.155)	-5.462 (4.855)	0.118
		2. Students like learning mathematics			
Grade 4	Main	-0.018 (0.047)	-0.067 (0.087)	0.028 (0.039)	0.317
	Placebo	0.032 (0.079)	0.050 (0.152)	0.015 (0.059)	0.826
Grade 8	Main	0.314*** (0.046)	0.346*** (0.068)	0.277*** (0.063)	0.459
	Placebo	0.007 (0.055)	0.026 (0.099)	-0.016 (0.039)	0.693
		3. Students confident in mathematics			
Grade 4	Main	0.028 (0.053)	-0.038 (0.091)	0.091* (0.054)	0.226
	Placebo	0.069 (0.057)	0.110 (0.095)	0.030 (0.069)	0.494
Grade 8	Main	0.046 (0.051)	0.009 (0.070)	0.089 (0.074)	0.431
	Placebo	0.021 (0.068)	0.019 (0.116)	0.022 (0.064)	0.981

*Notes: "Poorer − richer" reports the p-value for the difference between subgroup estimates. Placebo models use 2019 as a fictitious treatment period. * p<0.10, ** p<0.05, *** p<0.01.*

Supplementary Appendix C. DiD and Trend Estimates of Continuous School Closure Effects for Non-cognitive Factors.

Grade	Model	All countries	Poorer countries	Richer countries
		1. Students like learning mathematics		
Grade 4	DiD	0.008*** (0.003)	0.007 (0.006)	0.010*** (0.003)
	DiD Placebo	0.002 (0.003)	0.000 (0.007)	0.001 (0.002)
	Trend	0.005** (0.003)	0.010* (0.006)	0.005*** (0.002)
	Trend Placebo	-0.001 (0.006)	0.013 (0.008)	-0.010*** (0.003)
Grade 8	DiD	0.005 (0.005)	0.010 (0.006)	0.006 (0.006)
	DiD Placebo	-0.000 (0.003)	0.006 (0.004)	-0.010*** (0.003)
	Trend	0.008** (0.003)	0.005 (0.004)	0.019*** (0.005)
	Trend Placebo	0.007 (0.004)	0.009 (0.007)	0.001 (0.004)
		2. Students confident in mathematics		
Grade 4	DiD	-0.006* (0.003)	-0.007 (0.006)	-0.003 (0.002)
	DiD Placebo	-0.004 (0.004)	0.002 (0.006)	-0.007*** (0.002)
	Trend	-0.001 (0.005)	-0.006 (0.009)	0.005** (0.003)
	Trend Placebo	0.002 (0.006)	0.010 (0.009)	-0.004 (0.003)
Grade 8	DiD	0.001 (0.003)	0.001 (0.004)	0.008 (0.006)
	DiD Placebo	0.001 (0.002)	0.003 (0.003)	-0.006 (0.004)
	Trend	0.003 (0.003)	0.002 (0.004)	0.017** (0.007)
	Trend Placebo	0.009 (0.006)	0.015* (0.008)	0.002 (0.007)

Notes: Treatment is continuous and measured as weeks of full school closure. The "DiD" model estimates Difference-in-Differences effects, while the "Trend" model estimates deviations from country-level trends. Placebo models use 2019 as a fictitious treatment period. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Supplementary Appendix D. Parallel-trends tests for continuous-treatment DiD.

Grade	Model	Group	N	PTA p-value
		1. Mathematics		
Grade 4	DiD	All countries	740207	0.001***
	DiD	Poorer countries	321602	0.011**
	DiD	Richer countries	418605	0.000***
	DiD placebo	All countries	544556	0.197
	DiD placebo	Poorer countries	235157	0.007***
	DiD placebo	Richer countries	309399	0.001***
Grade 8	DiD	All countries	776514	0.007***
	DiD	Poorer countries	398595	0.020**
	DiD	Richer countries	377919	0.008***
	DiD placebo	All countries	569294	0.070*
	DiD placebo	Poorer countries	293240	0.035**
	DiD placebo	Richer countries	276054	0.001***
		2. Students like learning mathematics		
Grade 4	DiD	All countries	567494	0.773
	DiD	Poorer countries	257477	0.287
	DiD	Richer countries	310017	0.004***
	DiD placebo	All countries	378207	0.641
	DiD placebo	Poorer countries	174789	0.181
	DiD placebo	Richer countries	203418	0.001***
Grade 8	DiD	All countries	536464	0.030**
	DiD	Poorer countries	282759	0.379
	DiD	Richer countries	253705	0.033**
	DiD placebo	All countries	343177	0.009***
	DiD placebo	Poorer countries	182707	0.415
	DiD placebo	Richer countries	160470	0.038**

Grade	Model	Group	N	PTA p-value
		3. Students confident in mathematics		
Grade 4	DiD	All countries	566239	0.300
	DiD	Poorer countries	256615	0.484
	DiD	Richer countries	309624	0.003***
	DiD placebo	All countries	377680	0.216
	DiD placebo	Poorer countries	174345	0.301
	DiD placebo	Richer countries	203335	0.340
Grade 8	DiD	All countries	535226	0.276
	DiD	Poorer countries	282179	0.212
	DiD	Richer countries	253047	0.312
	DiD placebo	All countries	342672	0.120
	DiD placebo	Poorer countries	182463	0.096*
	DiD placebo	Richer countries	160209	0.435

*Notes: The table reports p-values from generalized parallel-trends tests for the continuous-treatment difference-in-differences (DiD) models. Statistical significance indicates that the parallel-trends assumption is violated. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Supplementary Appendix E. Robustness checks using alternative GDP per capita thresholds.

Grade	Model	Poorer 40th	Richer 40th	Poorer 60th	Richer 60th
1. Mathematics					
Grade 4	DiD	1.510* (0.836)	0.545*** (0.156)	0.614*** (0.213)	1.262 (0.784)
	DiD placebo	0.738 (0.863)	0.671*** (0.253)	0.608** (0.294)	1.379* (0.829)
	Trend	0.032 (0.796)	-0.097 (0.312)	-0.033 (0.321)	-0.764 (0.901)
	Trend placebo	-0.591 (1.097)	0.356 (0.238)	0.430* (0.230)	-0.785 (0.985)
Grade 8	DiD	-1.208 (0.971)	0.306 (0.199)	-0.213 (0.474)	1.155 (0.810)
	DiD placebo	-0.266 (0.585)	0.540** (0.249)	0.169 (0.270)	1.224* (0.734)
	Trend	0.227 (0.533)	-0.454 (0.315)	-0.134 (0.231)	-1.372* (0.795)
	Trend placebo	2.143*** (0.473)	-0.190 (0.193)	0.591 (0.454)	-1.806** (0.821)
2. Students like learning mathematics					
Grade 4	DiD	-0.007 (0.012)	0.010*** (0.002)	0.008** (0.003)	0.013 (0.009)
	DiD placebo	-0.019* (0.010)	0.005* (0.003)	0.001 (0.003)	0.002 (0.004)
	Trend	0.019 (0.017)	0.004** (0.002)	0.008** (0.003)	0.005 (0.006)
	Trend placebo	0.006 (0.024)	-0.002 (0.006)	0.003 (0.008)	-0.016** (0.006)
Grade 8	DiD	-0.003 (0.007)	0.013*** (0.003)	0.010 (0.006)	0.006 (0.006)
	DiD placebo	-0.000 (0.006)	0.003 (0.003)	0.006 (0.004)	-0.010*** (0.003)
	Trend	-0.003 (0.006)	0.013*** (0.002)	0.005 (0.004)	0.019*** (0.005)
	Trend placebo	-0.002 (0.008)	0.011*** (0.003)	0.009 (0.007)	0.001 (0.004)

Grade	Model	Poorer 40th	Richer 40th	Poorer 60th	Richer 60th
		3. Students confident in mathematics			
Grade 4	DiD	-0.005 (0.011)	-0.006* (0.003)	-0.006 (0.005)	-0.004 (0.005)
	DiD placebo	-0.016* (0.008)	-0.001 (0.004)	-0.002 (0.004)	-0.011** (0.005)
	Trend	0.013 (0.019)	-0.003 (0.006)	-0.002 (0.007)	0.010 (0.007)
	Trend placebo	-0.010 (0.018)	0.004 (0.006)	0.004 (0.007)	-0.004 (0.010)
Grade 8	DiD	-0.002 (0.007)	0.003 (0.003)	0.001 (0.004)	0.008 (0.006)
	DiD placebo	-0.000 (0.004)	0.003 (0.003)	0.003 (0.003)	-0.006 (0.004)
	Trend	0.000 (0.008)	0.004 (0.004)	0.002 (0.004)	0.017** (0.007)
	Trend placebo	0.005 (0.013)	0.014*** (0.004)	0.015* (0.008)	0.002 (0.007)

*Notes: The table presents robustness checks using alternative definitions of poorer and richer countries based on the 40th and 60th percentiles of GDP per capita. The "DiD" model estimates Difference-in-Differences effects, while the "Trend" model estimates deviations from country-level trends. Placebo models use 2019 as a fictitious treatment period. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Supplementary Appendix F. Robustness checks by country income group: excluding Kuwait and capping school closure duration at 50 weeks.

Grade	Specification	Model	All countries	Poorer countries	Richer countries
			1. Mathematics		
Grade 4	Without KWT	DiD	0.680** (0.305)	0.595 (0.414)	1.163 (0.804)
	Without KWT	DiD placebo	0.495 (0.369)	0.335 (0.462)	1.299 (0.837)
	Without KWT	Trend	0.105 (0.426)	0.145 (0.576)	-0.729 (0.882)
	Without KWT	Trend placebo	0.129 (0.431)	0.276 (0.562)	-0.686 (0.968)
	Capped	DiD	0.688*** (0.225)	0.595 (0.414)	0.863*** (0.312)
	Capped	DiD placebo	0.639** (0.305)	0.335 (0.462)	1.095*** (0.294)
	Capped	Trend	0.006 (0.327)	0.145 (0.576)	-0.393 (0.365)
	Capped	Trend placebo	0.292 (0.302)	0.276 (0.562)	0.146 (0.495)
Grade 8	Without KWT	DiD	-0.213 (0.494)	-0.436 (0.720)	1.150 (0.784)
	Without KWT	DiD placebo	0.310 (0.285)	0.219 (0.405)	1.236* (0.688)
	Without KWT	Trend	-0.338 (0.290)	-0.200 (0.361)	-1.338* (0.780)
	Without KWT	Trend placebo	0.490 (0.362)	0.926 (0.598)	-1.779** (0.826)
	Capped	DiD	-0.136 (0.406)	-0.436 (0.720)	0.463 (0.374)
	Capped	DiD placebo	0.292 (0.232)	0.219 (0.405)	0.570* (0.341)
	Capped	Trend	-0.296 (0.237)	-0.200 (0.361)	-0.562 (0.397)
	Capped	Trend placebo	0.404 (0.315)	0.926 (0.598)	-0.628 (0.543)
			2. Students like learning mathematics		
Grade	Without KWT	DiD	0.007 (0.005)	0.007 (0.006)	0.013 (0.009)

Grade	Specification	Model	All countries	Poorer countries	Richer countries
4	Without KWT	DiD placebo	0.002 (0.005)	0.000 (0.007)	0.002 (0.004)
	Without KWT	Trend	0.005 (0.004)	0.010* (0.006)	0.005 (0.006)
	Without KWT	Trend placebo	0.003 (0.008)	0.013 (0.008)	-0.016** (0.007)
	Capped	DiD	0.009** (0.004)	0.007 (0.006)	0.012*** (0.003)
	Capped	DiD placebo	0.002 (0.003)	0.000 (0.007)	0.002 (0.002)
	Capped	Trend	0.006* (0.003)	0.010* (0.006)	0.006*** (0.002)
	Capped	Trend placebo	-0.000 (0.007)	0.013 (0.008)	-0.012*** (0.003)
		3. Students confident in mathematics			
Grade 4	Without KWT	DiD	-0.008** (0.004)	-0.007 (0.006)	-0.003 (0.006)
	Without KWT	DiD placebo	-0.002 (0.005)	0.002 (0.006)	-0.011** (0.005)
	Without KWT	Trend	-0.004 (0.007)	-0.006 (0.009)	0.011* (0.006)
	Without KWT	Trend placebo	0.005 (0.008)	0.010 (0.009)	-0.004 (0.010)
	Capped	DiD	-0.007** (0.003)	-0.007 (0.006)	-0.004 (0.003)
	Capped	DiD placebo	-0.003 (0.004)	0.002 (0.006)	-0.009*** (0.002)
	Capped	Trend	-0.001 (0.006)	-0.006 (0.009)	0.007** (0.003)
	Capped	Trend placebo	0.003 (0.006)	0.010 (0.009)	-0.004 (0.004)

*Notes: The table reports robustness checks for the main results under two alternative specifications: (1) excluding Kuwait (KWT), which experienced exceptionally long school closures, and (2) capping school closure duration at 50 weeks to reduce the influence of Kuwait, which exceeds this threshold. Grade 8 non-cognitive outcomes are not reported because Kuwait is excluded from these analyses due to insufficient TIMSS cycle coverage. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl
ISSN 2957-0506