School closure policies and student reading achievement: evidence across countries

Alec I. Kennedy[1] · Rolf Strietholt[1]

Abstract

The COVID-19 pandemic disrupted education worldwide as educational systems made the decision to close schools to contain the spread of the virus. The duration of school closures varied greatly internationally. In this study, we use international variation in school closure policies to examine the effects of school closures on student achievement. Specifically, we use representative trend data from more than 300,000 students in 29 countries to examine whether the length of school closures is related to changes in student achievement before and after the outbreak of COVID-19. We observe a significant and substantial negative effect of school closures on student reading achievement. This school closure effect remains even after controlling for measures of pandemic severity such as infection rates, vaccination policies, and a measure of lockdown stringency. The estimated effect implies that a year of school closures corresponds roughly to the loss of a little more than half a school year of learning. This effect is even more pronounced for socioeconomically disadvantaged students and those without home computer access.

Keywords COVID-19 · Reading achievement · School closures · Elementary schools · International comparisons

1 Introduction

The COVID-19 pandemic caused widespread disruptions in education, with school closures affecting over 90% of students worldwide (UNESCO, 2020). The consequences of these closures on the learning and lives of children and their families have

✉ Rolf Strietholt
rolf.strietholt@iea-hamburg.de

Alec I. Kennedy
alec.kennedy@iea-hamburg.de

[1] Research and Analysis Unit, International Association for the Evaluation of Educational Achievement (IEA), Überseering 27, 22297 Hamburg, Germany

been the subject of public and scholarly concern. In particular, many speculate that school closures and disruptions may have affected learning outcomes. In the present study, we examine the relationship between the duration of national school closure policies and changes in achievement before and after the outbreak of the pandemic from an international perspective. Our goal is to better understand the effects of school closures on student learning.

An emerging body of literature has documented significant declines in student academic performance following the onset of the COVID-19 pandemic, with varying levels of reported learning deficits experienced across countries (e.g., Betthäuser et al., 2023; Di Pietro, 2023). However, the factors driving these across-country differences remain understudied. This knowledge gap arises from a limitation of prior research, as previous studies examining the impact of COVID-19 on education have relied mainly on national data. As most policies in response to the pandemic were implemented at the national level, there is often little variation within individual countries to evaluate the effects of these policies using national-level data. As a result, the link between school closure policies and learning declines across countries remains unclear. This paper seeks to address this gap by analyzing data from an international large-scale student assessment, which provides an internationally comparable measure of student reading achievement at the end of primary school.

Our study employs an international comparative approach to examine the impact of school closures on student learning (Strietholt & Scherer, 2018; Strietholt et al., 2014). The Progress in International Reading Literacy Study (PIRLS, Mullis et al., 2023) provides a unique opportunity for this type of research as it is the first international assessment to be administered after the onset of the pandemic. PIRLS collects internationally comparable data on fourth-grade reading achievement, along with important context data about classroom, home, and school learning environments. We supplement PIRLS data with information collected by UNESCO Institute for Statistics (UIS) on the duration and type of school closures implemented across several countries. With this combined data, we investigate the relationship between the duration of COVID-19-related national school closure policies and average reading performance across countries, while controlling for average achievement prior to the COVID-19 pandemic. Specifically, this study asks what impact did pandemic-related school closures have on trends in student reading achievement and how did it vary by length of school closure? We also explore how these effects vary by student background.

2 Background

In response to the COVID-19 pandemic, governments around the world implemented various measures to slow transmission of the virus, including school closures. The closure of schools presented a significant challenge for educational institutions, educators, students, and parents alike. The transition to remote or blended learning necessitated teachers to supply digital educational materials and monitor student progress. Moreover, the success of this approach depended on both teachers and students having access to the internet, the necessary equipment, suitable software, and apt Information and Communication Technology (ICT) skills (e.g., Strietholt et al., 2021; Stancel-Piatak

et al., 2023). School closures as a policy response were debated within countries as decision-makers weighed the benefits of school closures in containing the spread of the virus against the potential consequences to student learning. In the next section, we review the literature on school closure policies and student learning.

2.1 School closures policies during the COVID-19 pandemic

Over 1.5 billion students globally were impacted by decisions to shut down schools (OECD, 2021b). In this study, we focus on school closures spurred by the decision of policymakers to suspend or shut down in-person classes and activities in schools in response to the COVID-19 pandemic. The primary aim of these closures was to slow the transmission of the virus and safeguard the health of students, teachers, and staff. These decisions were mainly driven by assumptions extrapolated from the evidence of influenza outbreaks, which suggested that reduced social contacts between students could reduce virus transmission (e.g., Jackson et al., 2016; Viner et al., 2020). Furthermore, early correlational evidence also supported the idea that public health measures, including school closures, could be effective in slowing down the spread of the novel coronavirus (e.g., Auger et al., 2020; Pan et al., 2020).

While the benefits to school closure policies in battling against a global pandemic were highlighted in decisions to close schools, they were weighed against the potential costs to student learning and well-being. Much public and academic discourse has revolved around the potential consequences of school closures on the learning and lives of children and their families (e.g., Di Pietro et al., 2020; Meinck et al., 2022; Huber & Helm, 2020). Physical school closures likely impacted student learning in a variety of ways. First, moving from regular face-to-face instruction to remote learning practices limited student interaction with their teacher and peers reducing instructional time and student motivation to learn in some cases (Di Pietro et al., 2020). Second, teachers were also affected by these decisions as they were forced to adjust and teach outside of their regular classroom, sometimes using tools that they were unfamiliar with. This possibly impacted their ability to deliver instruction delaying the learning of their students (Rožman et al., 2022). Third, missing out on the socialization opportunities provided at school could have impacted student well-being and ultimately their learning progress (Rožman et al., 2022). Finally, home learning environments were also impacted by the pandemic. Families struggled with both the economic uncertainties brought about by governments' decisions to close businesses as well as the childcare concerns with their children no longer in school. Additional stress in the home, especially for those students from socioeconomically disadvantaged backgrounds, likely made it difficult for students to focus on learning (Rožman et al., 2022; Strietholt & Süttmann, 2022). As a result, there are serious concerns that school closures affected student learning progress during the pandemic.

2.2 International variation in duration of school closures

UNESCO (2020) has reported how school closures varied over time and across countries. Initially, schools were fully closed in most countries around the world as

educational systems and national governments sought to learn more about the spread of the new virus. However, as the pandemic progressed and more was learned, additional waves of school closures were determined mainly based on local case rates, with a general trend towards partially closed schools (e.g., with some grade levels being sent home or certain regions within a country experiencing closures) or fully closed schools with remote learning alternatives. Despite the global trend towards school closures, there were also large regional differences. For example, in the second half of 2021, most European countries had fully opened schools, while partial or full school closures were still dominant in Asian countries. This suggests that the impact of the pandemic on education varied across regions, with some areas experiencing greater disruption than others. According to data collected from over 200 countries by UIS, the length of school closures (full or partial) varied between 0 and 93 weeks (UNESCO, 2022).

Different reasons have been proposed to explain variation in school closure policies. The decision-making process regarding school closures was influenced by institutional systems and political orientations. For example, democratic countries tended to implement school closures quicker than those under more authoritarian regimes (Cronert, 2022). In addition, countries with higher governmental effectiveness tended to take longer than those with less effective state apparatuses (Cronert, 2022; Harris & Oliver, 2021). In decentralized countries, such as the USA, regional variation in school closure policies was more likely to occur, even within states. Furthermore, school closures were often tied closely with national or local case rates showing that health concerns motivated these decisions (Lindblad et al., 2021; Harris & Oliver, 2021). Another factor that likely played a role was the capacity of countries to offer quality alternatives to in-person instruction. Evidence shows that the ability of countries to offer quality remote learning varied both within- and across-countries (Barron Rodriguez et al., 2021; Muñoz-Najar et al., 2021; Kennedy et al., 2022).

2.3 Academic consequences of the COVID-19 pandemic

The learning deficit during COVID-19 refers to the academic setbacks students experienced after the outbreak of the pandemic. A growing number of studies are devoted to quantifying the learning deficit during the pandemic by comparing either the performance level or gains of students or student cohorts before and after the onset of the pandemic. Two recently published comprehensive and methodologically rigorous meta-analyses have synthesized the findings of this research on learning deficits, which have utilized national data from diverse countries (Betthäuser et al., 2023; Di Pietro, 2023). Drawing on over 45 individual studies from 18 countries, both meta-analyses revealed a significant decline in academic achievement following the outbreak of the pandemic, with similar mean effect sizes of Cohen's $d = -0.14$ and -0.17. These effect sizes are equivalent to approximately one-third to half a year's worth of learning. It is worth noting that there is substantial overlap in the studies analyzed in both meta-analyses. The findings are consistent with previous narrative reviews and meta-analyses that synthesized studies published shortly after the emergence of the COVID-19 pandemic, which predominantly reported on learning deficits (Donnelly

& Patrinos, 2021; Storey & Zhang, 2021; König & Frey, 2022; Patrinos & Vegas, 2022).[1]

Substantial variation in learning deficits can be observed across individual studies and countries, as demonstrated by two recent meta-analyses (Betthäuser et al., 2023; Di Pietro, 2023). The differences in learning deficits between countries are striking, and further research is needed to understand why some countries have experienced greater deficits than others. However, it is also important to recognize that the comparability of individual studies is limited due to several factors. These include differences in student populations, assessment domains, and assessment instruments used, as well as variation in the timing of the studies during the pandemic. Therefore, it is crucial to consider these limitations when interpreting the results of the meta-analyses, and general conclusions should be drawn with caution. At this point, the variability in estimated impacts has been understudied.

2.4 Change in educational inequality

The COVID-19 pandemic raised concerns among scholars and the public regarding the potential for an increase in the social achievement gap (Bailey et al., 2021; Goudeau et al., 2021). Various mechanisms were hypothesized on how school closures or educational disruptions could widen the achievement gap. For instance, parents in low-income families were more likely to be frontline workers, thereby exposing them to the virus, while also being less likely to have access to high-quality healthcare. Additionally, children from disadvantaged families may have had limited access to the digital resources that were essential for distance learning, resulting in a digital divide. Low-income parents may also have been less likely to provide academic support to their children, as they had fewer resources to supplement schooling with private tutoring. Moreover, such children were more likely to attend low-quality schools.

Research on learning deficits has examined whether the effects differ between socioeconomically disadvantaged and privileged children, but there is inconsistency in the findings. According to Betthäuser et al. (2023), pooling effect estimates from different studies is difficult due to varying indicators of socioeconomic background. Two-thirds of the studies considered in their meta-analysis reported a significant increase in social achievement inequality, while one-third found no significant changes. This finding is consistent with prior literature reviews (Donnelly & Patrinos, 2021; Hammerstein et al., 2021), but the reasons for inconsistency in the findings across studies remain unclear.

---

[1] We acknowledge that we do not cite all individual studies on the academic impacts of COVID-19. That would lead to an extensive list of studies. We invite readers interested in the extensive work done at the national level to read the numerous review articles cited here and identify studies of interest. With that said, several recent studies were brought to our attention which have yet to be included in systematic reviews or meta-analyses (Gambi & De Witte, 2023; Miller et al., 2023).

2.5 Effects of school closures on educational outcomes

To date, few studies have explored how the impacts of the pandemic on learning are associated with variation in school closure policies mainly because they have all been conducted at the national level where there is often little variation in school closure policies. One exception is in the USA where school closure decisions were often made at the local level and only sometimes guided by national guidelines. Jack et al. (2022) study this district-level variation in 11 states and find that school districts with full in-person learning (i.e., no school closures or hybrid learning) had significantly smaller declines in standardized test pass rates than those utilizing remote learning strategies. Moreover, the study suggests that the impact of school closures varies by social back-ground, with greater learning deficits observed in schools where a large proportion of students are Black or Hispanic. In addition, Patrinos (2023) uses estimated COVID-19 effects collected in a review of learning loss studies and relates it with the length of school closures. He uses lockdown stringency and vaccination measures as instrumental variables to estimate the causal impact of the length of school closures on student achievement. He estimates that a week of school closures leads to a decline of almost 1% of a standard deviation.

Beyond these studies, we are not aware of any other research that examines variation in COVID-19 effects on learning by differences in school closure policies. Our study seeks to fill this gap by providing some of the first evidence of the relationship between student learning progress and international variation in school closure policies, specifically focusing on the length of time schools were closed. In contrast to Patrinos (2023) who also estimates this relationship, we use internationally comparable data on reading achievement as opposed to information collected across several studies measuring achievement from different domains.

3 Data

The present study combines data from two sources to examine the relationship between school closure policies and reading achievement. The first data source is the Progress in International Reading Literacy Study (PIRLS), which provides internationally comparable data on reading achievement from five cycles spanning from 2001 to 2021. PIRLS measures the reading achievement of fourth-grade students in several countries around the world, making it a valuable resource for exploring the relationship between school closures and learning outcomes.

The second data source used in this study is information on school closure policies collected by the UNESCO Institute for Statistics (UIS). This data includes information on the length of school closures during the COVID-19 pandemic. By combining these two sources of data, we can examine how the length and type of school closures are associated with changes in reading achievement across countries.

To ensure the accuracy and reliability of the data, both PIRLS and UIS follow strict protocols for data collection and analysis. PIRLS uses standardized tests and carefully selected samples of students to ensure that the data collected is representative of

the population being studied. Similarly, UIS collects data from national education ministries and other official sources to ensure that the data is accurate and up-to-date.

3.1 Measures

3.1.1 Reading achievement

The outcome variable is reading achievement. The PIRLS assessment measures reading literacy through a range of tasks that assess comprehension, interpretation, and evaluation of literary and informational texts. PIRLS uses a rotated booklet and plausible value methodology to estimate students' achievement (Mislevy et al., 1992). All analyses presented below are based on estimation accounting for the variation across the five plausible values (Rubin, 2004). The achievement scores were transformed to a scale with an international mean of 500 and a standard deviation of 100 during the first cycle (PIRLS 2001). All subsequent cycles have been linked to this initial scale to allow the tracking of trends over time.

3.1.2 Duration of school closures

Our study employs the duration of school closures as the explanatory variable. The data on school closure duration across countries were obtained from UIS and are publicly available online (https://en.unesco.org/covid19/educationresponsehttps://en.unesco.org/covid19/educationresponse). Between February 2020 and March 2022, daily data was collected on the status of schooling systems across countries to monitor the extent and duration of school closures. For each country included in the database, daily data are available categorizing educational systems into four groups:

1. **Fully closed**: Government-mandates require schools to be closed affecting most or all students.
2. **Partially closed**: Schools are closed only in certain regions or for some grade levels. This also captures schools that are only partially open to in-person instruction (e.g., hybrid learning).
3. **Fully open**: Schools are open for face-to-face instruction for most or all students.
4. **Academic break**: Schools are on scheduled academic breaks for most or all students.

Our measure of school closure duration is constructed based on this information, accounting for both fully and partially closed days. Although partial school closures may have some impact, it is unclear from the data whether students in the target population of PIRLS are affected by these decisions to the same extent as fully closed days. Therefore, we construct a measure that weights fully closed days and partially closed days by a factor of 1 and 0.5, respectively. However, our findings remain consistent with different weighting methods.

The period over which we calculate the length of school closures depends on the timing of the data collection (see Table 1). We count the number of days in which schools were closed prior to the end of each data collection period. This means that

Table 1 List of countries in full sample by PIRLS 2021 data collection wave and school closure duration

Country	Cycles of data	Days of school closure
Wave 1 countries (assessed on schedule)		
Southern Hemisphere data collection dates: October–November 2020		
Northern Hemisphere data collection dates: March–June 2021		
Austria	2006, 2011, 2016, 2021	137
Azerbaijan, Republic of	2011, 2016, 2021	194
Belgium (Flemish)	2006, 2016, 2021	88
Belgium (French)	2006, 2011, 2016, 2021	88
Bulgaria	2001, 2006, 2011, 2016, 2021	149
Czech Republic	2001, 2011, 2016, 2021	163
Denmark	2006, 2011, 2016, 2021	107
Egypt	2016, 2021	91
Finland	2011, 2016, 2021	104
France	2001, 2006, 2011, 2016, 2021	47
Germany	2001, 2006, 2011, 2016, 2021	130
Italy	2001, 2006, 2011, 2016, 2021	128
Netherlands	2001, 2006, 2011, 2016, 2021	108
New Zealand	2001, 2006, 2011, 2016, 2021	27
Norway	2016, 2021	84
Oman	2011, 2016, 2021	118
Poland	2016, 2021	170
Portugal	2011, 2016, 2021	90
Russian Federation	2001, 2006, 2011, 2016, 2021	33
Singapore	2001, 2006, 2011, 2016, 2021	29
Slovak Republic	2001, 2006, 2011, 2016, 2021	121
Slovenia	2001, 2006, 2011, 2016, 2021	170
Spain	2006, 2011, 2016, 2021	63
Sweden	2001, 2006, 2011, 2016, 2021	61
Wave 3 Countries (assessed one year later)		
Southern Hemisphere data collection dates: August–December 2021		
Northern Hemisphere data collection dates: March–July 2022		
Australia	2011, 2016, 2021	111
England	2001, 2006, 2011, 2016, 2021	104
Iran, Islamic Republic of	2001, 2006, 2011, 2016, 2021	250
Israel	2011, 2016, 2021	121
South Africa	2016, 2021	194

Note: Countries included in the full sample are those with reading achievement in 2016 and 2021 and data on school closure duration. School closure duration has been rounded up to the nearest day

countries who participated in later data collection wave have potentially longer exposures to school closures based on our measure.

3.1.3 Student background

In addition to reading assessments, PIRLS also collects background information through context questionnaires. Given that we are examining trends in average student reading performance, we believe it is important to account for changes in the student population due to migration or economic developments that occur within a country. Therefore, we include several control variables in our analysis measuring characteristics of students. First, we include a measure of gender which, although it likely does not change much across cycles, can be important given the documented gender gap in reading present in many countries (e.g., Mullis et al., 2017, 2023). We also include age at the time of the test which can control for differences in testing time across cycles (Strietholt et al., 2013).

The next set of variables include responses to several items from the student questionnaire that were administered across all cycles of PIRLS. First, a question to students asking how many books they have in their home (five responses: 1 = 0–10; 2 = 11–25; 3 = 26–100; 4 = 101–200; 5 = More than 200) has been used to account for socioeconomic background (Mullis et al., 2017). Second, we also include a response about how often they speak the language of the test at the home (1 = Always or Almost Always; 2 = Sometimes; 3 = Never). Third, we also include student reports on whether they have access to a computer in the home. This variable can be important given the importance of technology during the time of school closures. When information was missing on any of these variables they were coded as a separate missing response as to not lose data from countries who did not administer these specific items for specific cycles. Descriptive statistics for the analytical sample as well as information on missing values can be found in Table 4 in the appendix.

Finally, in analyses using only more recent cycles of PIRLS data (e.g., 2016 and 2021), we use a composite measure of socioeconomic status (the Home Resources for Learning scale), which accounts for the number of books, children's books, and study supports in the home as well as the education and occupation of parents. This measure provides a more nuanced understanding of the socioeconomic context (e.g., Strietholt & Strello, 2022; Engzell, 2021; Jerrim & Micklewright, 2014 discuss measurement issues with using the books variable). Using item-response theory, a continuous scale is constructed with a centerpoint of 10 located at the mean of the combined distribution with units chosen so that 2 scale score points correspond to the standard deviation of the distribution (Martin et al., 2017).

3.2 Sample

Our study aims to examine the impact of the COVID-19 pandemic on fourth-grade students' reading performance. To accomplish this, we use data from PIRLS 2016 and PIRLS 2021 assessments in 29 countries that participated in both rounds of testing. Due to the pandemic, the administration of the PIRLS 2021 assessment was complex.

Twenty-four countries administered the assessment to fourth graders as usual at the end of the school year while five countries postponed the administration by one year and tested fourth graders 1 year later. We considered both sets of data to be useful. However, in some countries, fifth graders were tested at the beginning or middle of the school year instead of fourth graders at the end of the school year, making the data from these countries not directly comparable over time. Thus, following the PIRLS study center, we have removed these countries from our analysis to ensure the validity of our findings (Mullis et al., 2023). In addition to the 29 countries included in our analyses, there were three more countries—Hong Kong, Macao, and Taiwan—that administered the test at the end of grade four. However, the UIS database does not provide data on school closure policies for these countries. As a result, we were unable to include them in our analysis.

To establish a pre-pandemic baseline measure, we also add data from earlier rounds of PIRLS. However, since five countries did not participate in earlier cycles, our sample was restricted to 24 countries. Details on the countries, available data cycles, data collection wave for 2021, data collection timing for PIRLS 2021, and school closure duration are provided in Table 1.

The PIRLS assessment samples approximately 4000 students from 150 classes per country and per study cycle. To ensure that the sample of tested students is representative of the population of fourth-grade students in each participating country, a complex random sample design is employed. All analyses presented apply sampling weights. In pooled analyses, senate weights are used so that each country contributes equally to the final estimates, regardless of the sample and target population size. The standard errors are calculated using a jackknife repeated replication technique that accounts for sampling variance arising from the stratified class-based sampling design (Martin et al., 2017).

4 Methods

To examine the association between the duration of school closures ($Close_c$) and student reading achievement for country c (Y_c), we analyze pooled international data, comparing student performance during the pandemic with performance prior to the pandemic.

To control for baseline achievement levels that vary across countries, we use multiple approaches in our models. In the first approach, we only consider the average student achievement in 2016 as a benchmark, which shows how school closure duration relates to changes in average student reading achievement between the 2016 and 2021 PIRLS cycles. However, this approach overlooks possible longer-term trends in average student reading achievement observed in previous PIRLS cycles. To address this, we adopt a second model that includes all available cycles of PIRLS data for countries that participated in 2016 and 2021, estimating a global trend in average reading achievement. But this approach disregards the possibility of country-specific trends. Therefore, in our final model, we control for country-specific trends and focus on how the PIRLS 2021 results deviate from those trends.

With the inclusion of multiple cycles of data, we are able to control for country-specific time-invariant factors that likely influence both student achievement and their decision to close schools by including country-fixed effects. The country-fixed effects change the model to make within-country comparisons, effectively controlling for time-invariant country effects.

To ensure comparability between the samples of different years, it is essential to acknowledge potential changes in student composition, such as those arising from immigration. To account for these, we include several control variables. Specifically, our models include gender, the number of books in the home, how often students speak the language of the test at home, and whether they have a computer at home. Furthermore, age is included in the models to control for the time of testing that may differ across cycles.

With all of these considerations, we estimate the following model:

$$Y_{ict} = \alpha + \beta \text{Close}_c * I(t = 21) + \mu_c + f(t) + \gamma X_{ict} + \varepsilon_{ict} \tag{1}$$

where Y_{ict} is reading achievement for student i in country c during cycle t. It is important to note that we do not have repeated observations for students across cycles, so student i would not be observed in multiple t. That is, we do not have student panel data but country-level trend data. Close_c measures the duration of school closures in country c and is interacted with an indicator variable, $I(t = 21)$, so that the number of days of school closure is only considered in relation to PIRLS 2021 achievement results. μ_c represents country-fixed effects. $f(t)$ represents some functional form of time.

As mentioned above, we account for baseline achievement levels in three different models. In the first model, $f(t)$ is only a time-fixed effect for data with just two cycles (2016, 2021). In the second model, we fit a linear global trend in average student achievement. That is, $f(t) = \tau * time$ where $time$ is a continuous variable representing PIRLS cycles (1 = 2001, 2 = 2006, 3 = 2011, 4 = 2016, 5 = 2021) and τ is the estimated slope. In this model, we estimate a linear trend using the cycles conducted up to 2016 and project scores for 2021. Essentially, we compare the projected trend for 2021 with the actual scores observed in 2021. In the third model, we allow τ to vary by country, fitting country-specific trends in PIRLS average reading achievement: τ_c. We present estimates from each of these models.

Finally, X_{ict} represents a vector of student-specific control variables that includes measures of age, gender, home language, socioeconomic status (number of books in the home), and technology access. ε_{ict} is the error term that accounts for sampling variance due to the complex sampling design. β represents our parameter of interest. It is the estimated effect of an additional day of school closures on trends in average fourth-grade reading achievement.

To answer our research question regarding the heterogeneity of the effects by student background, we modify equation 1 by replacing $\text{Close}_c * I(t = 21)$ with a triple-interaction term. We interact the number of days of school closure and the 2021 indicator with several student characteristics: a measure of student socioeconomic status, whether or not they report having a computer in the home, language background, and gender. Furthermore, to account for different trends for these groups of students,

we interact our functional form of time with these student background characteristics. We estimate each model separately for each student's background characteristics. All relevant main effects are also included in the models.

5 Results

We begin by presenting descriptive patterns in the data. Figure 1 presents the bivariate relationship between school closure duration and changes in average reading achievement observed between PIRLS 2016 and PIRLS 2021. The scatterplot shows a general negative relationship between school closure duration and changes in average reading achievement. That is, countries that closed schools for longer periods of time, based on our measure, tended to show larger declines in average reading performance than those countries that closed schools for shorter periods of time.

A similar exercise can be done where country-specific slopes are estimated which we present in the Appendix in Figs. 2 and 3. Results produce similar findings: a negative relationship between school closure duration and reading achievement.

5.1 School closure duration and change in reading achievement

Table 2 contains the estimated effects of one day of school closures on average student reading achievement as measured on the PIRLS scale obtained from our three model
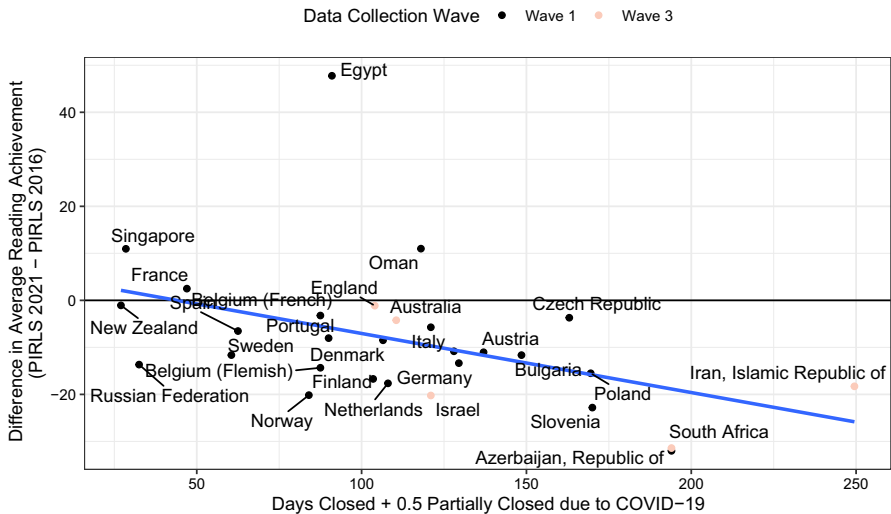

Fig. 1 Relationship between school closure duration and changes in average student reading achievement between PIRLS 2016 and 2021. Note: The correlation between the two measures is moderate ($r = -0.44$). Egypt stands out as an outlier in the plot as it showed by far the largest positive change in average reading achievement. However, there are some questions about the reliability of the results in Egypt as between 15 and 25% of students had achievement too low for estimation (Mullis et al., 2023). Therefore, we urge the reader to interpret the results in Egypt with caution. Removing Egypt from the sample leads to a slightly stronger correlation ($r = -0.56$)

	Δ16–21	Global trend	Country trends
Table 2 Estimated effect of one day of school closures on average student reading achievement			
Estimate	−0.140***	−0.174***	−0.118***
(S.E.)	(0.015)	(0.012)	(0.018)
# Cty	29	24	24
# Cty x Yr	58	103	103
# Student	327,639	527,995	527,995

*p<0.05; **p<0.01; ***p<0.01

Note: Δ16-21 shows estimates from a model of the change between PIRLS 2016 and PIRLS 2021. Global trend shows estimates from a model that accounts for global trends in average student reading achievement. Country trends shows results from a model that accounts for country-specific trends in average student reading achievement. All regressions have utilized sampling weights (senate weights). Standard errors have been calculated accounting for sampling variation using jackknife repeated replication

specifications. The first column shows the results from the model that focuses just on the change between PIRLS 2016 and PIRLS 2021 (Δ16-21). The second column shows the results from a model that accounts for a global trend in average student reading performance (Global Trend). The final column presents the estimate from the model that specifically models country-specific trends in reading achievement (Country Trends). All models include student-level controls for age, gender, number of books in the home, home language, and whether the student has a computer in the home. The models also include country-fixed effects and controls for the PIRLS 2021 data collection wave.

In all models, the estimates tell a similar story: longer school closures are significantly associated with larger declines in average student reading performance. The effect sizes range from −0.12 in the Country Trends model to −0.14 in the Δ16-21 model to −0.17 in the Global Trend model. All coefficients are in PIRLS scale point units. While the magnitude of each estimate varies, the only two coefficients that are significantly different from each other are the Country Trends and Global Trend model estimates. With this in mind, we choose to focus on the result from the PIRLS 2016 to PIRLS 2021 specification (i.e., −0.14) as it allows for a larger sample of countries while also providing an estimate that is not significantly different from the other two methods. It also provides an average estimate across the three models.

5.2 School closure and change in inequalities in achievement

We next test the heterogeneity of the school closure duration coefficient across student characteristics. For these results, we focus on findings from the Δ16-21 model. This allows us to examine the heterogeneity across more characteristics as it only uses data from the PIRLS 2016 and 2021 cycles. Namely, we are able to use the Home Resources for Learning scale that was available during these two cycles. We opt to use this measure in an interaction as opposed to the measure of the number of books

in the home as it incorporates information on both the number of books in the home as well as parental education and occupation. We summarize the findings in Table 3.

The first rows shows the estimates of the interaction between the school closures measure and the continuous home resources for learning scale. We centered the scale to the international average (10), so that the coefficient estimate on the school closure measure (without the interaction) can be interpreted as the effect at the average level of home resources. The coefficient on the school closure measure is very similar to what is obtained in the averaged results (-0.15). The coefficient on the interaction term shows how the estimated effect of school closures changes as students' values on the home resources for learning scale moves. The coefficient is significant and positive (0.021) suggesting that the effect is less negative for students with higher values on the home resources for learning scale and that it is more negative for students with lower values.

Table 3 Heterogeneity in the effect of one day of school closures on average student reading achievement

Variable	Estimate	Standard error
Home Resources for Learning Scale (continuous)		
Closed	−0.146***	(0.014)
Closed x Home Resources for Learning	0.021*	(0.008)
# Cty	28	
# Cty x Yr	56	
# Student	256,929	
Do you have a computer at home? (categorical)		
Yes (baseline)	−0.143***	(0.014)
No	−0.085*	(0.041)
How often do you speak the test language at home? (categorical)		
Always or Almost Always (baseline)	−0.137***	(0.017)
Sometimes	−0.022	(0.029)
Never	0.031	(0.036)
Gender (categorical)		
Girls (baseline)	−0.148***	(0.016)
Boys	0.015	(0.022)
# Cty	29	
# Cty x Yr	58	
# Student	327,639	

*$p<0.05$; **$p<0.01$; ***$p<0.01$

Note: All heterogeneity models use the Δ16-21 model. Results are shown from separate models that interact our measure of school closures with: the continuous (1) Home Resources for Learning scale centered at the international average (10) and the categorical (2) Do you have a computer at home?, (3) How often do you speak the test language at home? and (4) Gender. For the continuous interaction, we report the main effect (Closed) and interaction effect (Closed x Home Resources for Learning). For the categorical interactions, we report coefficients for the baseline category (highlighted with a (baseline) next to it) as well as differences from the baseline category. Note that one country (England) was dropped from the analysis on Home Resources for Learning as they did not administer the home questionnaire that was used to construct the scale. All regressions have utilized sampling weights (senate weights). Standard errors have been calculated accounting for sampling variation using jackknife repeated replication

These findings suggest that there is evidence that social inequalities were exacerbated by school closure policies. The Home Resources for Learning scale exhibits a standard deviation of 2. Based on the analyses, this indicates that the impact of one day of school closures on a student with a socioeconomic status one standard deviation below the average is estimated to be -0.186 ($-0.146 - 2 * 0.021$). Conversely, for a student with a socioeconomic status one standard deviation above the average, the estimated impact is -0.104 ($-0.146 + 2 * 0.021$).

The next rows report inequalities across student groupings. These are interactions with categorical variables, so the coefficient tests whether the group coefficient estimate differs significantly from the baseline group (which is identified in the table). In examining how the effect of school closures differs for students with and without a computer, we find that students without a computer in the home had a significantly more negative effect of school closures (-0.09 more negative) than those with a computer. In examining how the effect differs for language or gender, we do not find significant differences in the estimated school closure effects for each group.

5.3 Robustness tests

Several extended tests confirm the robustness of our preferred model specification where we combined 2016 and 2021 PIRLS data from 29 countries and used the weighted number of full and partial school closure days (weighted as 1 and 0.5, respectively) as the main explanatory variable (see Appendix).

First, while we attempt to control for confounding factors in our analysis, there might be concerns that what our estimates are capturing is a general COVID effect rather than the effect of school closures. For instance, the severity of the pandemic in some countries is likely correlated with the length of school closures while also affecting student achievement through all the ways the pandemic impacted daily lives. To attempt to control for this and isolate the effect of school closures, we run our main model including controls for country-level case rates and death rates per capita collected by the World Health Organization (WHO) (https://covid19.who.int/WHO-COVID-19-global-table-data.csv). The data used is the cumulative confirmed cases and deaths per 100,000 population to date. In addition, we include data collected from the COVID-19 Government Response Tracker (OxCGRT) (https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker). Specifically, we use their stringency index measure which accounts for information on whether governments implemented several policies restricting movement or interaction in response to the pandemic (e.g., school closures, workplace closings, cancellation of public events, restrictions on gatherings, rules for public transportation, stay at home orders, restrictions on internal movement, international travel controls, or public information campaigns). In addition, we use information on the number of days before the majority of the population was vaccinated as well as the vaccination rate at the time of data collection. The length of school closures is positively correlated with deaths per capita ($r = 0.24$), vaccination percentage at date of data collection ($r = 0.32$), number of days until the majority of the population was vaccinated ($r = 0.32$), and the stringency index ($r = 0.37$). In contrast, the correlation with cases per capita is

negative ($r = -0.26$). The inclusion of these five measures as controls into our main model does not change our main takeaways (-0.15 versus -0.14). In addition, similar to Patrinos (2023), we use these measures as instruments, and estimate an effect of school closures as -0.15 (see first two columns of Table 5 for these results).

Second, there might be concerns that non-participation may have biased the results of this study (e.g., Werner & Woessmann, 2023). On average, weighted participation rates declined about five percentage points between the PIRLS 2016 and PIRLS 2021, ranging from a 12 percentage point drop in Oman to a three percentage point increase in Portugal. Exclusion rates, which measures the percentage of schools and students from the target population that were excluded prior to sampling, did not change much across cycles (on average, about a half a percentage point increase). With these concerns in mind, we include participation and exclusion rates as control variables in our models to ensure that our estimated effects are not capturing any correlation between school closures and non-participation in PIRLS (see final column of Table 5). After controlling for this information, our estimate changes slightly, but the story remains the same (-0.10 versus -0.14). It should be noted that a decline in participation would likely lead to an underestimation of the school closure effect given that non-participation is more likely among the students hardest hit by the pandemic (Werner & Woessmann, 2023).

Third, the administration of PIRLS 2021 was postponed by one year in five countries. Consequently, during the outbreak of COVID-19, the children in these countries were one grade lower than the children in the other countries. Furthermore, students in these countries have potentially longer exposures to school closures prior to assessment. In order to examine whether this has an impact on our findings, we replicated our main analyses excluding these five countries. The results of the main analysis are qualitatively the same as they are for all 29 countries (the second column in Table 6).

Fourth, it is important to note that our primary explanatory variable is a national measure of school closures. However, in decentralized countries, there can be variations in school closure policies at regional or local levels. To address this issue, we excluded eight federal countries and replicated the main analyses. Again, the results remained consistent (third column in Table 6).

Fifth, the UIS measurements of partial school closures are not as detailed as we may want for this type of analysis. In our main analyses, we pragmatically assigned a weight of 0.5 to this category trying to account for the fact that students assessed in PIRLS may or may not be impacted by these partial school closures. In subsequent analyses, we explored alternative weighting approaches on the partial school closure measure, using weights of 0 (disregarding partial closures) and 1 (considering partial closures equivalent to full closures). Using alternative measures of our main explanatory variable school closure revealed some impact on our analyses but the main result remained robust across these different weighting methods, as shown in Table 7.

Finally, it is widely recognized that outliers can have a significant impact on regression analyses. In Fig. 1, Egypt stands out as a potential outlier. To examine whether our results are heavily influenced by a single country's inclusion in the analysis, we performed multiple re-estimations of our main model, systematically excluding a different country each time. The results consistently demonstrated a high degree of similarity, indicating that our main finding is not reliant on any specific country (see Fig. 4).

6 Conclusion

While an expanding body of literature has noted significant learning deficits following the onset of COVID-19, the explanation for why these deficits vary remains unclear. To shed light on this issue, we utilize data from PIRLS 2021—one of the first international assessments conducted after the onset of the COVID-19 pandemic—to investigate the relationship between national school closure policies and changes in fourth-grade average academic achievement at an international level.

6.1 School closures are linked to declines in achievement and increased inequality

This study presents the first evidence of how international variation in school closure policies was related to student academic performance. We estimate that an additional day of school closures is associated with a 0.14 PIRLS scale point decline in student reading performance. On average, countries in our sample closed schools for 110 days. To put our estimated effect size into context, this would imply that, the average country's average student reading achievement declined about 15 points more than a country that did not close schools. With an international standard deviation of 100, this would be an average effect of 0.15 SD which aligns remarkably well with results from several meta-analyses (e.g., Storey & Zhang, 2021; König & Frey, 2022; Betthäuser et al., 2023; Di Pietro, 2023).

To provide further context for our findings, we can refer to previous PIRLS cycles, which assessed students from various grade levels in specific countries. By examining data from PIRLS 2016, where Denmark (grades 3 and 4) and Norway (grades 4 and 5) had students from multiple grade levels participating, we observed average score differences of 46 and 42 PIRLS points, respectively (Mullis et al., 2017). Assuming an approximate improvement of 44 points on the PIRLS scale over a typical school year (which includes both a schooling and maturation effect), we can utilize this information to gauge the learning loss resulting from school closures. Note, however, that the learning progress over one year might differ across countries (see Steinmann & Olsen, 2022). Considering that a standard school year comprises of around 180 days in various countries (OECD, 2021a), we can estimate that a full year of school closures would entail a loss slightly above half (57%) of a school year's worth of learning (calculation: 180 * 0.14/44 = 0.57).

Another main finding of the present study is that we observed evidence suggesting that social inequalities have been amplified due to school closure policies. Specifically, we observed that the effect of school closures was more pronounced for socioeconomically disadvantaged students and those without home computer access. To put our results into context, our estimates would indicate that the achievement gap between a student one standard deviation above the international average on the home resources for learning scale and a student one standard deviation below would grow by about 9 points in an average country (i.e., one that closed for 110 days). An increase of a similar magnitude would also be observed between students with and without a computer in the home. This indicates a disturbing likelihood that students who usually trail behind their peers academically may have fallen further behind due to the decisions to

shut down schools. These findings align with other studies investigating the impacts of COVID-19 on social inequalities. Recovery initiatives should take these observations into account, prioritizing support for students from less advantaged socioeconomic backgrounds. In contrast to socioeconomic background and home computer access, we do not observe any varying effects in relation to gender or the languages spoken by students at home.

Evidence showing that other subjects (i.e., mathematics) were impacted more by the pandemic than reading makes it important to note that the true impact of the school closure policies may be larger than the effects estimated in this study (Betthäuser et al., 2023; Di Pietro, 2023). As data from other international studies are released, it will be important to understand whether similar patterns can be observed across learning domains.

6.2 Limitations

The present study is subject to some limitations that warrant acknowledgment.

First, the implementation of PIRLS during the pandemic posed significant challenges, potentially impacting the quality and reliability of the collected data. While we attempted to address this concern by excluding countries that did not administer PIRLS 2021 in grade 4, it is important to acknowledge that other participating countries might still have been affected by data quality issues.

Second, the categorization of "partially closed" schools, one of the school closure categories from the UIS database, is not as detailed as one would hope for an analysis like this. Specifically, the measure does not indicate whether the PIRLS target population were affected by school closures. We attempted to address this concern by applying different weighting approaches to construct our main explanatory variable; however, a more in-depth analysis of different school closure policies might yield more insight into the effects. In the same vein, computer access was operationalized with a simple dichotomous measure.

Third, our measurement of school closures was conducted at the national level, which overlooks potential variations within countries. Although we partially addressed this limitation by excluding decentralized countries with federal states, there may still be within-country variations in school closure policies that are not captured well in the partial school closure measure.

Fourth, while we utilized international trend data, controlled for prior achievement, and incorporated a comprehensive set of student-level controls, it is important to acknowledge that country-level measures may be correlated with school closures, as governments may implement various policy packages to mitigate COVID-19 infections (as discussed by Goodman-Bacon & Marcus, 2020). This could introduce potential confounding factors in our analysis. It is worth noting that these parallel policies would only bias our findings if they are correlated with student achievement. For instance, measures such as mask mandates or vaccination policies may be associated with school closures but are unlikely to be directly linked to student achievement. One other concern might be that we are not necessarily capturing the impacts of school closures, but an overall COVID effect. We attempt to address part of this issue

by including country-level measures of COVID-19 health impacts (i.e., case rates per capita and death rates per capita) to separate out the different effects and do not observe a change in the overall estimates. Nonetheless, further research on this topic is necessary to thoroughly investigate potential correlations and their impact.

Fifth, it is important to note that the models used to analyze global and country-specific trends assume linearity. While the assumption of linearity may be subject to question, it is worth noting that our limited number of observations prevents us from accurately estimating quadratic or other non-linear trends with a fuller sample of countries. The consistency of our results across different model specifications makes us feel confident that our results are not fully a product of our model assumptions.

Sixth, it is important to note that our investigation primarily centers on reading proficiency at the end of primary school. Future studies might explore other areas like mathematics and other stages of education, including secondary school. Additionally, we recommend that any comprehensive evaluation of school closures should also consider process-related variables, such as the stress experienced by educators, students, and parents.

Despite the aforementioned limitations, we believe that our study provides valuable evidence shedding light on the consequences of school closures on student achievement during the pandemic. We are especially encouraged that the results are within a range consistent with other studies.

6.3 Educational and economic implications

The observed decline in student performance following school closures carries profound implications. Previous research indicating the stability of student achievement over time underscores the significance of the finding, suggesting that the learning deficits resulting from these closures can have lasting and long-term consequences for educational careers. As Hanushek and Woessmann (2020) highlight, these learning deficits may extend beyond individual educational outcomes and have economic ramifications for both the affected individuals and national economic growth. Recognizing the magnitude of these consequences, it is imperative for secondary and tertiary education sectors to be adequately prepared to address and mitigate the impact of these learning deficits. By proactively implementing strategies and interventions, secondary and tertiary institutions can play a pivotal role in minimizing the long-term effects and ensuring that students are equipped to navigate their educational journeys successfully in the aftermath of school closures.

Additionally, in terms of educational inequality, our findings underscore the importance of providing particular attention and support to disadvantaged children who may be disproportionately affected by the learning deficits resulting from school closures. By targeting interventions towards these vulnerable populations, we can strive towards equitable educational outcomes for all students. These results also highlight the importance of making sure these students have adequate learning resources in the home in the event that schools must be closed again.

As educational systems made the difficult choice to close schools in an effort to prevent the spread of COVID-19, there was limited evidence available on the benefits

and costs of such decisions. Results from this study can be combined with the extensive research on the benefits of school closure policies in slowing virus transmission to develop a more holistic understanding of the consequences of school closures. This knowledge may help educational systems make better-informed decisions in the event of any future disruptions to education due to a global pandemic or other events such as teacher strikes, natural disasters, or armed conflicts.

Appendix

Table 4 Descriptive statistics of analytical sample

Continuous variables

Variable	Obs	Mean	SD	Min	Max
Reading PV1	528,204	526.06	85.95	5.00	830.63
Reading PV2	528,204	525.45	86.51	7.54	836.16
Reading PV3	528,204	525.04	86.51	8.35	847.97
Reading PV4	528,204	525.25	86.60	5.00	898.40
Reading PV5	528,204	525.32	86.35	5.00	855.14
Age at Test	528,021	10.27	0.55	6.21	14.98

Categorical variables

	Category	Obs	Percent
Gender			
	Female	260,648	49.35
	Male	267,497	50.65
Books at home			
	0–10	69,679	13.19
	11–25	120,737	22.86
	26–100	170,842	32.34
	101–200	82,911	15.70
	200+	69,385	13.14
	Missing	14,650	2.77
How often do you speak the test language at home?			
	Always/Almost Always	341,316	64.62
	Sometimes	94,276	17.85
	Never	58,119	11.00
	Missing	34,493	6.53
Do you have a computer at home?			
	Yes	446,807	84.59
	No	70,636	13.37
	Missing	10,761	2.04

Note: Descriptive statistics are shown for the dataset used in the analysis which includes countries who participated in at least the 2011, 2016, and 2021 cycles of PIRLS as well as have data on school closure duration in the database. Missing variables were treated as a separate category. Summary statistics are across all countries and cycles in the sample. No sampling weights have been applied

Table 5 Estimated effect of one day of school closures on average student reading achievement with different control variables

Ctrl for	Pandemic Msrs	IV: Pandemic Msrs	Participation
Estimate	−0.153***	−0.149***	−0.103***
(S.E.)	(0.018)	(0.022)	(0.014)
# Cty	29	29	29
# Cty x Yr	58	58	58
# Student	327,639	327,639	327,639

*p<0.05; **p<0.01; ***p<0.01

Note: All results are from the model that examines the change between PIRLS 2016 and PIRLS 2021 (Δ16-21). The first column includes additional controls for weighted participation rates and exclusion rates. The second column includes pandemic measures as controls: overall case rates per capita, overall death rates per capita, vaccination rate at time of testing, number of days until the majority of the population was vaccinated, and a stringency index that measures the level of restrictions of government policies related to COVID-19. The final column uses the pandemic measures as instruments in an instrumental variables regression. All regressions have utilized sampling weights (senate weights). Standard errors have been calculated accounting for sampling variation using jackknife repeated replication
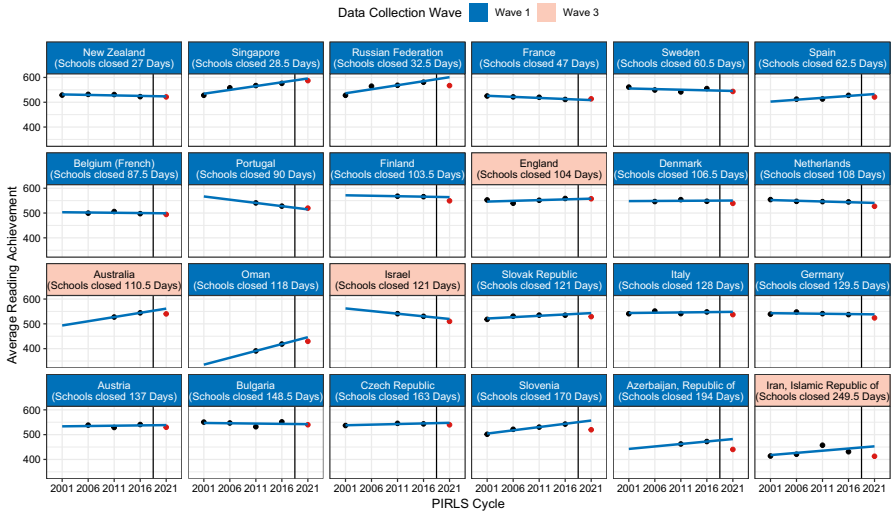
Table 6 Estimated effect of one day of school closures on average student reading achievement by different samples of countries

	Wave 1 only	Wave 1 + 3	No federal countries
Estimate	−0.136***	−0.140***	−0.144***
(S.E.)	(0.016)	(0.015)	(0.016)
# Cty	24	29	21
# Cty x Yr	48	58	42
# Student	262,253	327,639	220,878

*p<0.05; **p<0.01; ***p<0.01

Note: All results are from the model that examines the change between PIRLS 2016 and PIRLS 2021 (Δ16-21). The first column only includes countries who were part of the first data collection wave in PIRLS 2021 (i.e., those that collected data as regularly scheduled). The second column shows results from a sample that includes both Wave 1 and Wave 3 (those tested one year later). The third column presents results when removing federal decentralized countries (A list of federal decentralized countries was obtained here: https://forumfed.org/countries/https://forumfed.org/countries/). Specifically, excluding federal decentralized countries removes Australia, Austria, Belgium (Flemish and French), Germany, the Russian Federation, South Africa, and Spain from the regression. All regressions have utilized sampling weights (senate weights). Standard errors have been calculated accounting for sampling variation using jackknife repeated replication

Fig. 2 Country-specific trends in PIRLS average reading achievement. Note: Country-specific trends are estimated only for data points prior to the 2021 cycle and extrapolated
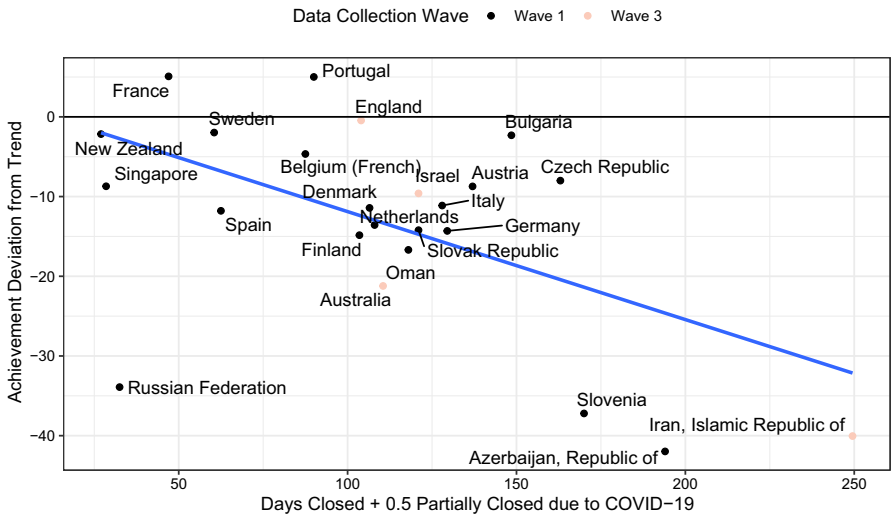

Fig. 3 Relationship between school closure duration and deviations from country-specific trends. Note: Achievement deviation from the trend is calculated as the difference between the extrapolated linear trend prior to PIRLS 2021 for a country and the actual observed point in PIRLS 2021. Figure 2 visualizes the country-specific extrapolation by country. The correlation is moderate ($r = -0.55$)

Table 7 Estimated effect of one day of school closures on average student reading achievement by different measures of school closures

	Full	Full + 0.5 partial	Full + partial
Estimate	−0.125***	−0.140***	−0.102***
(S.E.)	(0.018)	(0.015)	(0.010)
# Cty	29	29	29
# Cty x Yr	58	58	58
# Student	327,639	327,639	327,639

*p<0.05; **p<0.01; ***p<0.01

Note: All results are from the model that examines the change between PIRLS 2016 and PIRLS 2021 (Δ16-21). The first column only counts days in which schools were fully closed. The second column sums the days in which schools were fully closed and weights partial school closures by one half. The third column sums the days in which schools were either fully or partially closed. All regressions have utilized sampling weights (senate weights). Standard errors have been calculated accounting for sampling variation using jackknife repeated replication
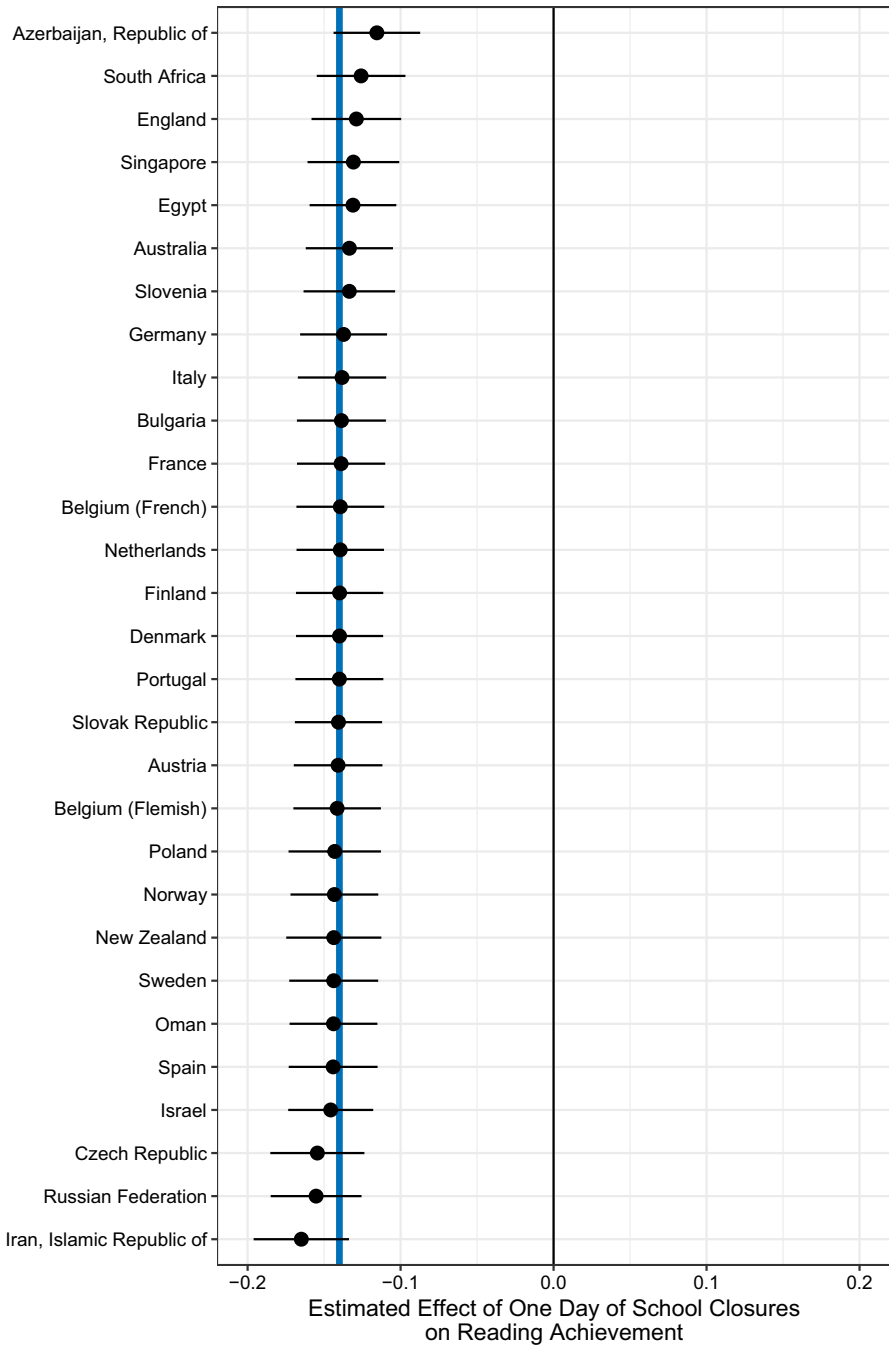
Fig. 4 Distribution of Estimated Effects of School-Closure Duration on Average Student Reading Achievement after Removing One Country. Note: Each point represents a coefficient estimate of school closures on average reading achievement after removing one country from the sample. The blue line represents the overall effect estimate. Line ranges represent 95% confidence intervals

Data Availability   The data that support the findings of this study are available from the corresponding author upon request.

Declarations

Competing interests   The authors declare no competing interests.

References

Auger, K. A., Shah, S. S., Richardson, T., Hartley, D., Hall, M., Warniment, A., et al. (2020). Association between statewide school closure and COVID-19 incidence and mortality in the US. *Jama, 324*(9), 859–870.

Bailey, D. H., Duncan, G. J., Murnane, R. J., & Au Yeung, N. (2021). Achievement gaps in the wake of COVID-19. *Educational Researcher, 50*(5), 266–275.

Barron Rodriguez, M., Cobo, C., Munoz-Najar, A., & Sánchez Ciarrusta, I. (2021). *Remote learning during the global school lockdown: Multi-country lessons*. World Bank: Remote Learning During the Global School Lockdown.

Betthäuser, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour, 7*(3), 375–385.

Cronert, A. (2022). Precaution and proportionality in pandemic politics: democracy, state capacity, and COVID-19-related school closures around the world. *Journal of Public Policy, 42*(4), 705–729. https://doi.org/10.1017/S0143814X22000101

Di Pietro, G. (2023). The impact of COVID-19 on student achievement: Evidence from a recent meta-analysis. Educational Research Review 100530

Di Pietro, G., Biagi, F., Costa, P., Karpiński, Z., & Mazza, J. (2020). The likely impact of COVID-19 on education: Reflections based on the existing literature and recent international datasets (Vol. 30275). Publications Office of the European Union Luxembourg

Donnelly, R., & Patrinos, H. A. (2021). Learning loss during COVID-19: An early systematic review. Prospects 1–9

Engzell, P. (2021). What do books in the home proxy for? A cautionary tale. *Sociological Methods & Research, 50*(4), 1487–1514.

Gambi, L., & De Witte, K. (2023). *The uphill battle: The amplifying effects of negative trends in test scores, COVID19 school closures and teacher shortages*. Department of Economics: KU Leuven.

Goodman-Bacon, A., & Marcus, J. (2020). Using difference-in-differences to identify causal effects of COVID-19 policies. DIW Berlin Discussion Papers

Goudeau, S., Sanrey, C., Stanczak, A., Manstead, A., & Darnon, C. (2021). Why lockdown and distance learning during the COVID-19 pandemic are likely to increase the social class achievement gap [Journal Article]. *Nat Hum Behav, 5*(10), 1273–1281. https://doi.org/10.1038/s41562-021-01212-7. https://www.ncbi.nlm.nih.gov/pubmed/34580440

Hammerstein, S., Konig, C., Dreisorner, T., & Frey, A. (2021). Effects of COVID-19-related school closures on student achievement-A systematic review [Journal Article]. *Front Psychol, 12*. https://doi.org/10.3389/fpsyg.2021.746289. https://www.ncbi.nlm.nih.gov/pubmed/34603162

Hanushek, E. A., & Woessmann, L. (2020). The economic impacts of learning losses. OECD

Harris, D. N., & Oliver, D. M. (2021). Why did so many public schools stay remote during the COVID crisis? National Center for Research on Education Access and Choice

Huber, S. G., & Helm, C. (2020). COVID-19 and schooling: Evaluation, assessment and accountability in times of crises-reacting quickly to explore key issues for policy, practice and research with the school barometer. *Educational assessment, evaluation and accountability, 32*, 237–270.

Jack, R., Halloran, C., Okun, J., & Oster, E. (2022). Pandemic schooling mode and student test scores: Evidence from US school districts. American Economic Review: Insights

Jackson, C., Vynnycky, E., & Mangtani, P. (2016). The relationship between school holidays and transmission of influenza in England and Wales. *American journal of epidemiology, 184*(9), 644–651.

Jerrim, J., & Micklewright, J. (2014). Socio-economic gradients in children cognitive skills: Are cross-country comparisons robust to who reports family background? *European Sociological Review, 30*(6), 766–781.

Kennedy, A. I., Mejía-Rodríguez, A. M., & Strello, A. (2022). Inequality in remote learning quality during COVID-19: Student perspectives and mitigating factors. *Large-scale Assessments in Education, 10*(1), 29.

König, C., & Frey, A. (2022). The impact of COVID-19-related school closures on student achievement-A meta-analysis. *Educational Measurement: Issues and Practice, 41*(1), 16–22.

Lindblad, S., Wärvik, G.-B., Berndtsson, I., Jodal, E.-B., Lindqvist, A., Messina Dahlberg, G., et al. (2021). School lockdown? Comparative analyses of responses to the COVID-19 pandemic in European countries. *European Educational Research Journal, 20*(5), 564–583.

Martin, M. O., Mullis, I. V., & Hooper, M. (2017). Methods and procedures in PIRLS 2016 (Tech. Rep.). Boston: Boston College, TIMSS & PIRLS International Study Center

Meinck, S., Fraillon, J., & Strietholt, R. (2022). The Impact of the COVID-19 Pandemic on Education: International Evidence from the Responses to Educational Disruption Survey (REDS). International Association for the Evaluation of Educational Achievement

Miller, A., Fray, L., & Gore, J. (2023). Was COVID-19 an unexpected catalyst for more equitable learning outcomes? A comparative analysis after two years of disrupted schooling in Australian primary schools. The Australian Educational Researcher 1–22

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29*(2), 133–161.

Mullis, I. V., Martin, M. O., Foy, P., & Hooper, M. (2017). PIRLS 2021 International Results In Reading (Tech. Rep.). Boston College, TIMSS & PIRLS International Study Center. Retrieved from https://timssandpirls.bc.edu/pirls2016/

Mullis, I. V., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., & Wry, E. (2023). PIRLS 2021 International Results in Reading (Tech. Rep.). Boston College, TIMSS & PIRLS International Study Center. Retrieved from https://timssandpirls.bc.edu/pirls2016/

Mullis, I. V., von Davier, M., Foy, P., & Fishbein, B. (2023). PIRLS 2021 International Results in Reading (Tech. Rep.). Boston College, TIMSS & PIRLS International Study Center. Retrieved from https://pirls2021.org/results

Muñoz-Najar, A., Gilberto, A., Hasan, A., Cobo, C., Azevedo, J. P., & Akmal, M. (2021). *Remote learning during COVID-19: Lessons from today*. World Bank, Washington, DC: Principles for Tomorrow.

OECD (2021a). Education at a glance 2021: OECD indicators (Tech. Rep.). Paris. Retrieved from https://doi.org/10.1787/b35a14e5-en

OECD (2021b). The state of school education: One year into the COVID pandemic (Tech. Rep.). Paris. Retrieved April 12, 2023 from https://doi.org/10.1787/201dde84-en

Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., et al. (2020). Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan. *China. Jama, 323*(19), 1915–1923.

Patrinos, H. A. (2023). The longer students were out of school, the less they learned. *Journal of School Choice, 17*(2), 161–175.

Patrinos, H. A., Vegas, E., & Carter-Rau, R. (2022). *An analysis of COVID-19 student learning loss*. Washington, DC: World Bank.

Rožman, M., Meinck, S., & Chen, M. (2022). Impact of the pandemic on classroom teaching and learning. In S. Meinck, J. Fraillon, & R. Strietholt (Eds.), *The impact of the COVID-19 pandemic on education:*

International evidence from the Responses to Educational Disruption Survey (REDS) (pp. 54–83). Amsterdam, Netherlands: UNESCO and IEA.

Rožman, M., Meinck, S., & Chen, M. (2022). Well-being of students and teachers. The impact of the COVID-19 pandemic on education: International evidence from the Responses to Educational Disruption Survey (REDS) (pp. 54–83). Amsterdam, Netherlands: UNESCO and IEA

Rubin, D. B. (2004). Multiple imputation for nonresponse in surveys (Vol. 81). Wiley

Stancel-Piatak, A., Liaw, Y.-L., Strietholt, R., & Koop, A.-K. (2023). Prepared for distance learning? International evidence on the use and perception of information and communication technology by teachers. Global Perspectives on Education Research, vol. II (pp. 11–33). Routledge

Steinmann, I., & Olsen, R. V. (2022). Equal opportunities for all? Analyzing within-country variation in school effectiveness. *Large-Scale Assessments in Education, 10*(1), 1–34.

Storey, N., & Zhang, Q. (2021). A meta-analysis of COVID learning loss. Preprint at EdArXiv, 1–15

Strietholt, R., Bos, W., Gustafsson, J.-E., & Rosén, M. (2014). Educational policy evaluation through international comparative assessments. Waxmann Verlag

Strietholt, R., Fraillon, J., Liaw, Y.-L., Meinck, S., & Wild, J. (2021). *Changes in digital learning during a pandemic-Findings from the ICILS teacher panel*. International Association for the Evaluation of Educational Achievement: International Association for the Evaluation of Educational Achievement.

Strietholt, R., Rosén, M., & Bos, W. (2013). A correction model for differences in the sample compositions: The degree of comparability as a function of age and schooling. *Large-scale Assessments in Education, 1*(1), 1–20.

Strietholt, R., & Scherer, R. (2018). The contribution of international large-scale assessments to educational research: Combining individual and institutional data sources. *Scandinavian Journal of Educational Research, 62*(3), 368–385.

Strietholt, R., & Strello, A. (2022). Socioeconomic inequality in achievement: Conceptual foundations and empirical measurement. In: International Handbook of Comparative Large-Scale Studies in Education: Perspectives, Methods and Findings (pp. 1–20). Springer

Strietholt, R., & Süttmann, F. (2022). Inequalities in teaching and learning during the pandemic. In S. Meinck, J. Fraillon, & R. Strietholt (Eds.), *The impact of the COVID-19 pandemic on education: International evidence from the Responses to Educational Disruption Survey (REDS)* (pp. 54–83). Amsterdam, Netherlands: UNESCO and IEA.

UNESCO (2020). School closures caused by coronavirus (COVID-19). Retrieved from March 10, 2022. https://en.unesco.org/covid19/educationresponse

UNESCO (2022). UNESCO map on school closures and UIS. Retrieved from https://en.unesco.org/covid19/educationresponse(Accessed on September 1, 2022)

Viner, R. M., Russell, S. J., Croker, H., Packer, J., Ward, J., Stansfield, C., Mytton, O., Bonell, C., & Booy, R. (2020). School closure and management practices during coronavirus outbreaks including COVID-19: A rapid systematic review. *The Lancet Child & Adolescent Health, 4*(5), 397–404.

Werner, K., & Woessmann, L. (2023). The legacy of COVID-19 in education. Economic Policy eiad016