A systematic review and meta-analysis of the impact of the COVID-19 pandemic on learning

Bastian A. Betthäuser[1–4], Anders M. Bach-Mortensen[3], Per Engzell[5,2,4,6]

How has the COVID-19 pandemic affected learning progress among school-age children? A growing number of studies address this question, but findings vary depending on context. We conduct a pre-registered systematic review, quality appraisal and meta-analysis of 34 studies across 12 countries to assess the magnitude of the effect of the pandemic on learning. We find a substantial overall learning deficit early in the pandemic (Cohen's $d = -0.17$, 95% c.i. -0.22, -0.13), which persists over time. Forgone learning is particularly large among children from low socio-economic backgrounds, and in contexts with high excess mortality, longer school closures, and less-developed digital learning infrastructures. The geographic reach of the existing evidence is limited, as studies mainly focus on high-income countries. Future research should fill the evidence gap in middle- and low-income countries and avoid the common risks of bias that we identify.

The COVID-19 pandemic has led to one of the largest disruptions to learning in history. To a large extent this is due to school closures, which are estimated to have affected 95 percent of the world's student population.[1] But even when face-to-face teaching resumed, instruction has often been compromised by hybrid teaching, and by teachers or children having to quarantine and miss classes. The effect of limited face-to-face instruction is likely compounded by the pandemic's consequences for children's out-of-school learning environment, as well as their mental and physical health. Lockdowns have restricted children's movement and their ability to play, meet other children, and engage in extra-curricular activities. Children's well-being and family relationships have also suffered due to economic uncertainties and conflicting demands of work, care and learning. These negative consequences can be expected to be most pronounced for children from low socio-economic family backgrounds, exacerbating pre-existing educational inequalities.

It is critical to understand how the COVID-19 pandemic has affected children's learning progress. We use the term 'learning deficit' to encompass both a delay in expected learning progress, as well as a loss of skills and knowledge already gained. The COVID-19 learning deficit is likely to affect children's life chances through their education and labor market prospects. At the societal level, it can have important implications for growth, prosperity, and social cohesion. As policy-makers across the world are seeking to limit further learning deficits and to devise policies to recover learning deficits that have already been incurred, assessing the current state of learning is crucial. A careful assessment of the COVID-19 learning deficit is also necessary to weigh the true costs and benefits of school closures.

We conduct a systematic review and meta-analysis of the evidence on COVID-19 learning deficits two years into the pandemic. Our first contribution is to review the existing evidence and to appraise its geographic reach and quality. More specifically, we ask (a) How much evidence is there on the effect of the COVID-19 pandemic on learning?, (b) Which countries are represented in the available evidence?, and (c) What is the quality of the existing evidence?

Our second contribution is to harmonize, synthesize and meta-analyze the existing evidence, with special attention to variation across different sub-populations and country contexts. Based on the studies we identify, we ask (d) What is the magnitude of the overall effect of the COVID-19 pandemic on learning?, (e) How has the magnitude of the learning deficit evolved since the beginning of the pandemic?, (f) To what extent has the pandemic reinforced inequalities between children from different socio-economic backgrounds? (g) Are there differences in the magnitude of the learning deficit between subject domains (math and reading) and between grade levels (primary and secondary)?, and (h) To what extent do contextual factors, such as country context, excess mortality, length of school closure, and teachers' use of digital learning tools moderate the effect of the pandemic on learning progress? Below, we report our answers to each of these questions in turn.

The state of the evidence

Our systematic review identified 34 studies on the effect of the COVID-19 pandemic on learning. As shown in Fig. 1, the initial literature search resulted in 5,153 hits after removal of duplicates. We also hand-searched relevant preprint repositories and policy databases, which yielded 97 further studies. All studies were double-screened by the first two authors. Studies had to measure learning using standardized test scores to be eligible for inclusion. The initial search process identified 28 eligible studies. To ensure that our study selection was as up-to-date as possible, we conducted a full forward and backward citation search of all included studies on February 15, 2022. This allowed us to identify 6 additional studies, giving us a total of 34 studies. Most of these studies were published after the initial search, which illustrates that the body of evidence on the effect of the pandemic on learning continues to expand. Studies often provide multiple estimates of COVID-19 learning deficits, typically referring to different subjects (math and

[1]Observatoire Sociologique du Changement (OSC), Sciences Po
[2]Leverhulme Centre for Demographic Science, University of Oxford
[3]Department of Social Policy and Intervention, University of Oxford
[4]Nuffield College, University of Oxford
[5]Social Research Institute, University College London
[6]Swedish Institute for Social Research, Stockholm University

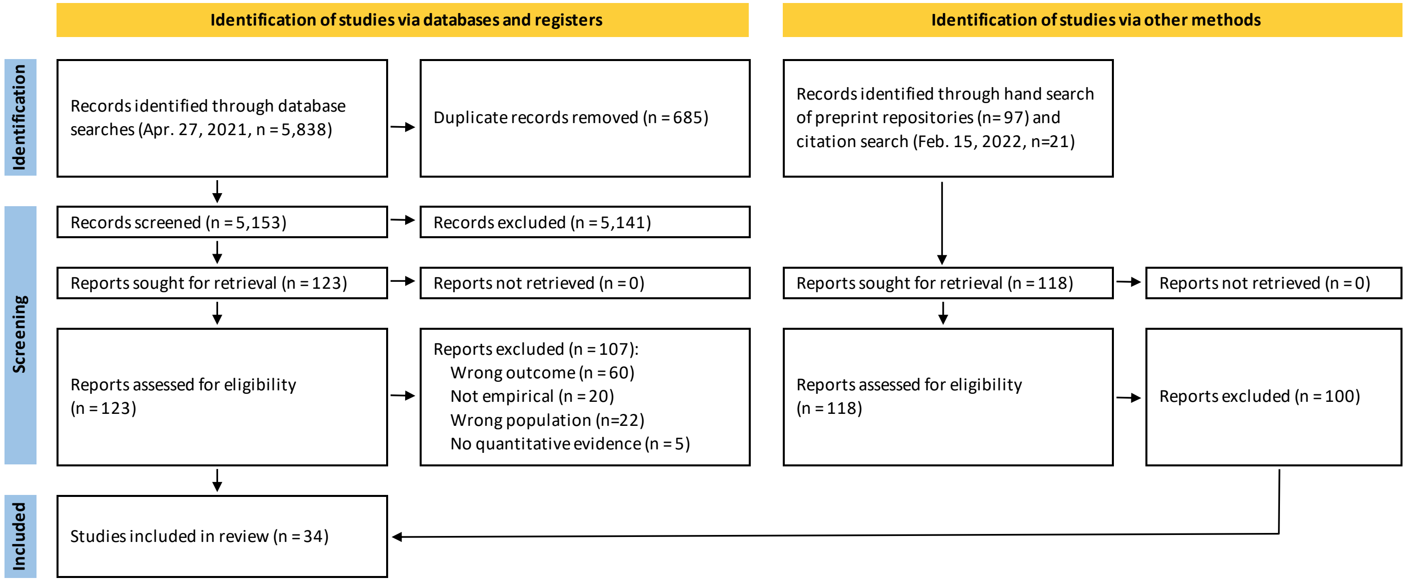


Figure 1: Study identification and selection process (PRISMA flow diagram)

reading) or school grades. The number of estimates (n=235) is therefore larger than the number of studies (n=34).

The geographic reach of the evidence is limited. Table 1 shows all identified studies and estimates of COVID-19 learning deficits (in brackets), grouped by country. 12 countries are represented: Australia, Belgium, Brazil, Denmark, England, Germany, Italy, Mexico, the Netherlands, South Africa, Switzerland, and the United States. About half of the estimates ($n = 95$) are from the United States, 64 are from the United Kingdom, a further 36 are from other European countries, and the remaining 10 estimates are from Australia, Brazil, Mexico, and South Africa. As this list shows, there is a strong over-representation of studies from high-income countries, a dearth of studies from middle-income countries, and no studies from low-income countries. This skewed representation should be kept in mind when interpreting our synthesis of the existing evidence on COVID-19 learning deficits.

The quality of evidence is mixed. We assessed the quality of the evidence using an adapted version of the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool.[34] More specifically, we analyzed the risk of bias of each estimate from confounding, sample selection, classification of treatments, missing data, the measurement of outcomes, and the selection of reported results. The second author performed the risk of bias assessments, which were independently checked by the first and third author. We then assigned each estimate an overall risk of bias rating (low, moderate, serious, or critical) based on the estimate and domain with the highest risk of bias. In line with ROBINS-I guidance, we excluded all studies rated 'critical' ($n = 2$) from our meta-analysis.[34]

Fig. 2*A* shows the distribution of all estimates of COVID-19 learning deficits according to their risk of bias rating separately for each domain (top six rows), as well as the distribution of estimates according to their overall risk of bias rating (bottom row). The overall risk of bias was considered 'critical' for 6% of estimates, 'serious' for 42% of estimates, 'moderate' for 44% of estimates, and 'low' for 8% of estimates. As shown in Fig. 2*A*, common sources of potential bias were confounding, sample selection, and missing data. The likely consequence of these sources of bias is an underestimation of COVID-19 learning deficits. Studies rated at risk of confounding typically compared only two time points, without accounting for longer time trends in learning progress. The main causes of selection bias were the use of convenience samples and/or insufficient consideration of self-selection by schools or students. Several studies found evidence of selection bias, often with students from a low socio-economic background or schools in deprived areas being underrepresented after (as compared to before) the pandemic, but this was not always adjusted for. Some studies also reported a higher amount of missing data post-pandemic, again generally without adjustment, and several studies did not report any information on missing data.

No evidence of publication bias. Publication bias can occur if authors self-censor to conform to theoretical expectations, or if journals favor statistically significant results. To mitigate this concern, we include not only published papers, but also unpublished working papers and 'gray literature', such as policy reports. Moreover, Fig. 2*B* tests for publication bias by showing the distribution of z-statistics for the effect size estimates of all identified studies. The dotted line indicates $z = 1.96$ ($p = 0.05$), the conventional threshold for statistical significance. The overlaid curve shows a normal distribution. If there was publication bias, we would expect a spike just above the threshold, and a slump just below it. There is no indication of this, and publication bias does thus not appear to be a major concern.

Results

Having assessed the quality of the existing evidence, we now present the substantive results of our meta-analysis, focus-

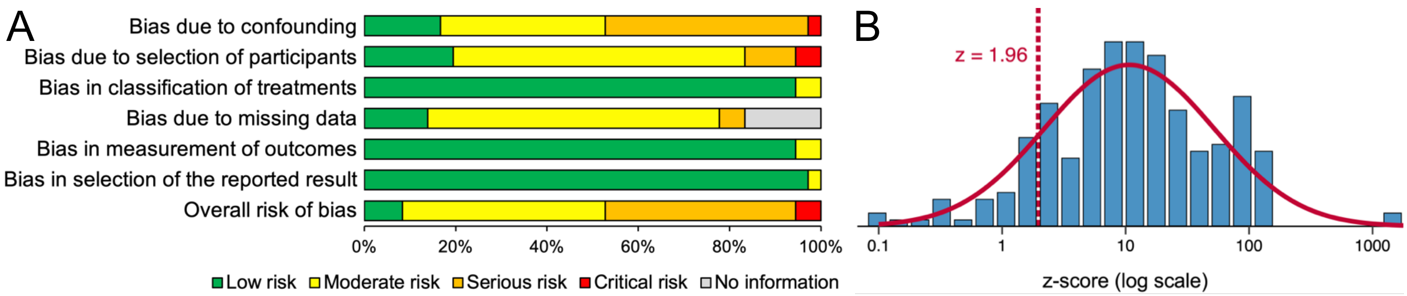


Figure 2: Risk of bias and publication bias. (*A*) Domain-specific and overall distribution of estimates of COVID-19 learning deficits (n=235), by risk of bias rating using ROBINS-I; (*B*) *z*-curve: Distribution of the z-scores of all estimates (log scale) to test for publication bias. The dotted line indicates $z = 1.96$ ($p = 0.05$), the conventional threshold for statistical significance. The overlaid curve shows a normal distribution. The absence of a spike in the distribution of the z-scores just above the threshold for statistical significance, and the absence of a slump just below it, indicate an absence of evidence for publication bias.

Table 1: Studies and estimates by country

Country	Studies
Australia [4]	Gore et al. 2021 [4][2]
Belgium [4]	Gambi and De Witte 2021 [2],[3] Maldonado and De Witte 2021 [2][4]
Brazil [2]	Lichand et al. 2021 [2][5]
Denmark [7]	Birkelund et al. 2021 [7][6]
Germany [6]	Depping et al. 2021 [4],[7] Schult et al. 2021 [2][8]
Italy [1]	Contini et al. 2021 [1][9]
Mexico [2]	Hevia et al. 2022 [2][10]
Netherlands [16]	Engzell et al. 2021 [8],[11] Haelermans et al. 2021 [2],[12] Schuurman et al. 2021 [6][13]
South Africa [2]	Ardington et al. 2021 [2][14]
Switzerland [2]	Tomasik et al. 2020 [2][15]
United Kingdom [64]	Blainey et al. 2020 [10],[16] Blainey et al. 2021a [12],[17] Blainey et al. 2021b [12],[18] Blainey et al. 2021c [12],[19] Department for Education 2021a [6],[20] Department for Education 2021b [2],[21] GL Assessment 2021 [4],[22] Rose et al. 2021a [2],[23] Rose et al. 2021b [4][24]
United States [95]	Domingue et al. 2021a [8],[25] Domingue et al. 2021b [4],[26] Kogan and Lavertu 2021a [1],[27] Kogan and Lavertu 2021b [9],[28] Kozakowski et al. 2021 [12],[29] Kuhfeld et al. 2020 [12],[30] Lewis et al. 2021a [12],[31] Lewis and Kuhfeld 2021b [12],[32] Pier et al. 2021 (MAP) [10], Pier et al. 2021 (Star) [10], Pier et al. 2021 (iReady) [5][33]

Note: Countries and corresponding studies on COVID-19 learning deficits. The number of estimates are shown in brackets, by country (left) and study (right). Full references are indicated by superscript and listed in the bibliography.

ing on the magnitude of COVID-19 learning deficits and on the variation in learning deficits over time, across different groups of students, and across different country contexts.

The COVID-19 pandemic led to substantial learning deficits. Fig. 3 shows the effect sizes that we extracted from each study (averaged across grades and learning subject) as well as the pooled effect size (red diamond). Effects are expressed in standard deviations, using Cohen's *d*. Estimates are pooled using inverse variance weights. The pooled effect size across all studies is $d = -0.17$. Under normal circumstances, students generally improve their performance by around 0.4 standard deviations per school year.[35,36,37] Thus, the overall effect of $d = -0.17$ suggests that students lost out on 0.17/0.4, or about 40%, of a school year's worth of learning. On average, the pandemic has led to a substantial learning deficit.

Learning deficits arise early in the pandemic and persist over time. One may expect that children were able to recover learning that was lost early in the pandemic, after teachers and families had time to adjust to the new learning conditions, and structures for online learning and for recovering early learning deficits were set up. However, existing research on teacher strikes in Belgium[38] and Argentina,[39] shortened school years in Germany,[40] and disruptions to education during World War II[41] suggests that learning deficits are difficult to compensate and tend to persist in the long run.

Fig. 4 plots the magnitude of estimated learning deficits (on the vertical axis) by the date of measurement (on the horizontal axis). The color of the circles reflects the relevant country, the size of the circles indicates the sample size for a given estimate, and the smoothed line displays a running average. The figure shows that learning deficits opened up early in the pandemic and have neither closed nor widened since then. This suggests that efforts by children, parents, teachers, and policy-makers to adjust to the changed circumstance have been successful in preventing further learning deficits, but so far have been unable to reverse them.

The pandemic exacerbated educational inequality. Existing research on the development of learning gaps during summer vacations,,[42,43] disruptions to schooling during the Ebola outbreak in Sierra Leone and Guinea,[44] and the

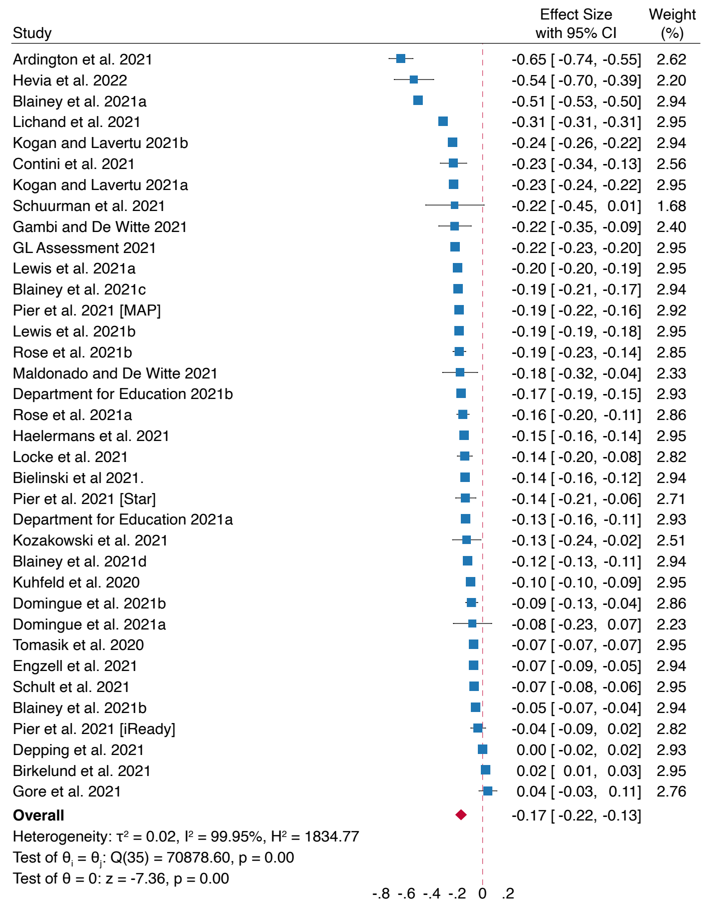


Figure 3: Forest plot showing individual estimates by study (averaged across subjects and grade levels), and the overall effect size estimate, pooled using inverse variance weights. Effect sizes are expressed in standard deviations, using Cohen's d, with 95% confidence intervals.

2005 earthquake in Pakistan,[45] shows that the suspension of face-to-face teaching can increase educational inequality between children from different socio-economic backgrounds. The effect of the COVID-19 pandemic on learning progress is likely to have been particularly pronounced for children from low socio-economic backgrounds. These children have been more affected by school closures than children from more advantaged backgrounds.[46] Moreover, they are likely to be disadvantaged with respect to their access and ability to use digital learning technology, the quality of their home learning environment, and the learning support they receive from teachers and parents.[47,48,49]

Most studies we identify examine the effect of the pandemic on socio-economic inequality, attesting to the importance of the issue. Because studies use different measures of socio-economic background (e.g., parental income, parental education, free school meal eligibility, neighborhood disadvantage), pooling the estimates is not possible. Instead, we code all estimates according to whether they indicate a positive, negative, or no effect of the pandemic on learning inequality. Fig. 5 displays this information. Estimates that indicate an increase in inequality are shown on the right, those that indicate a decrease on the left, and those that suggest no change in the middle. Squares represent estimates on the effect of the pandemic on inequality in reading performance, and circles represent estimates on the effect of the pandemic on inequality in math performance. The shading represent when in the pandemic educational inequality was measured: 0–6 months, 7–12 months, or 13–19 months after the onset of the pandemic in March 2020. Estimates are also arranged horizontally by grade level. A large majority of estimates indicate an increase in educational inequality between children from different socio-economic backgrounds. This holds for both math and reading, across primary and secondary education, at each stage of the pandemic, and independently of how socio-economic background was measured.

Learning deficits are larger in math than in reading. Available research on summer learning deficits,[42,50] student absenteeism,[51,52] and extreme weather events,[53] suggests that learning progress in mathematics is more dependent on formal instruction than in reading. This might be due to parents being better equipped to help their children with reading, and children advancing their reading skills (but not their math skills) when reading for enjoyment outside of school. Fig. 6*A* shows that similar to earlier disruptions to learning, the estimated COVID-19 learning deficits are larger for math (mean -0.192, median -0.190) than for reading (mean -0.130, median -0.115). This difference is statistically significant and robust to dropping estimates from individual countries, except the United States, which accounts for about half of all estimates (see Fig. S4).

Learning deficits do not vary across grade levels. One may expect learning deficits to be smaller for older than for younger children, as older children may be more autonomous in their learning and better able to cope with a sudden change in their learning environment. Fig. 6*B* shows that, contrary to expectation, we find no marked difference in the learning deficits between younger and older students. Note, however, that secondary students were subject to longer school closures in some countries, such as Denmark,[6] based partly on the assumption that they would be better able to learn from home. This may have offset any advantage that older children would otherwise have had in learning from home.

Learning deficits are larger in poorer countries. Low and middle-income countries were already struggling with a learning crisis before the pandemic. Despite large expansions of the proportion of children in school, children in low and middle-income countries still perform poorly by international standards, and inequality in learning remains high.[54,55,56] The pandemic is likely to deepen this learning crisis and to undo past progress. Schools in low- and middle-income countries have not only been closed for longer, but have also had fewer resources to facilitate remote learning.[57,58] Moreover, the economic resources, ICT equipment and ability of children, parents, teachers, and governments to support learning from home are likely to be lower in low- and middle-income countries.[59]

As discussed above, most evidence on COVID-19 learning deficits comes from high-income countries. We found no studies on low-income countries that met our inclusion

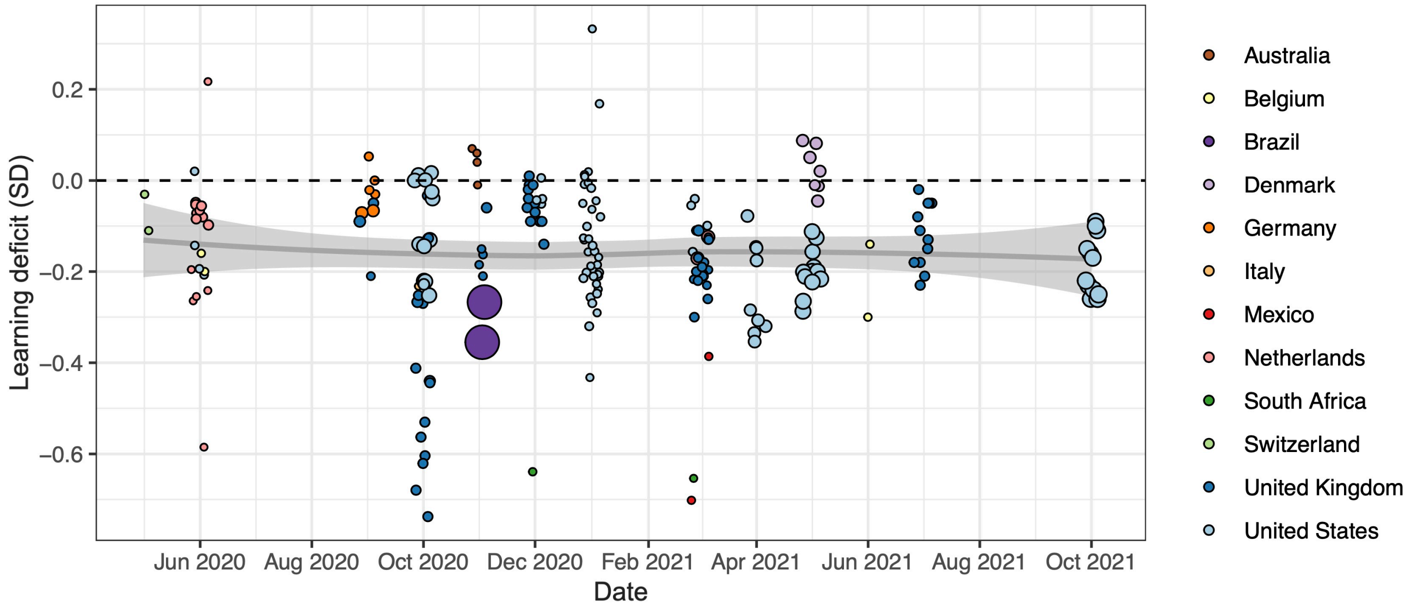


Figure 4: Estimates of COVID-19 learning deficits (n=235), by date of measurement. The horizontal axis displays the date on which learning progress was measured. The vertical axis displays estimated learning deficits, expressed in standard deviations using Cohen's d. The color of the circles reflects the respective country, the size of the circles indicates the sample size for a given estimate, and the smoothed line displays a running average with a 95% confidence interval.

criteria, and evidence from middle-income countries is limited to Brazil, Mexico and South Africa. Fig. 6*C* groups the estimates of COVID-19 learning deficits in these three middle-income countries together (on the right) and compares them to estimates from high-income countries (on the left). The learning deficit is appreciably larger in middle-income countries (mean −0.500, median −0.513) than in high-income countries (mean −0.149, median −0.143). In fact, the 3 studies on middle-income countries in our sample[14,10,5] are among the 4 studies reporting the largest estimates of learning deficits in Fig. 3.

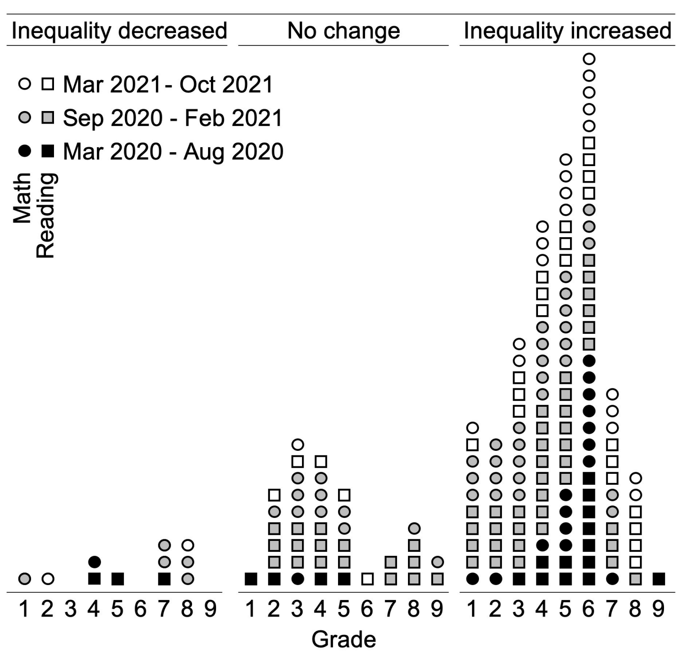


Figure 5: Harvest plot summarizing the evidence on the effect of the pandemic on educational inequality between students from different socio-economic backgrounds. Each circle/square refers to one estimate of over-time change in inequality in math/reading performance. Estimates that find a decrease/no change/increase in inequality are grouped on the left/middle/right. Within these categories, estimates are ordered horizontally by grade level. The shading indicates when in the pandemic a given measure was taken.

Higher excess mortality is associated with larger learning deficits. High incidences of infection and mortality due to the pandemic have been shown to cause not only physical stain but also psychological distress in young people.[60] In turn, this is likely to affect their learning progress. Moreover, a more severe course of the pandemic is associated with more drastic measures to contain the virus, including (but not limited to) school closures, hybrid learning, curfews and teleworking, which constitute additional obstacles to learning.

We measure the severity of the pandemic as the cumulative excess mortality at the country level. This measure is preferable over case numbers or registered COVID-19 deaths, which are difficult to compare across countries. Fig. 6*D* plots all estimates of COVID-19 learning deficits (on the vertical axis) against excess mortality (on the horizontal axis) in the relevant country and at the date when learning deficits where measured. It shows a clear pattern, where learning deficits are higher in countries that have experienced a more severe course of the pandemic.

Longer school closures are associated with larger learning deficits. The pandemic has presented various challenges to students' learning, but the largest of these has arguably been the suspension of face-to-face teaching, which is estimated to have affected 95 per cent of the world's student population.[57,58] Remote learning may be a poor substitute for face-to-face learning. This is true particu-

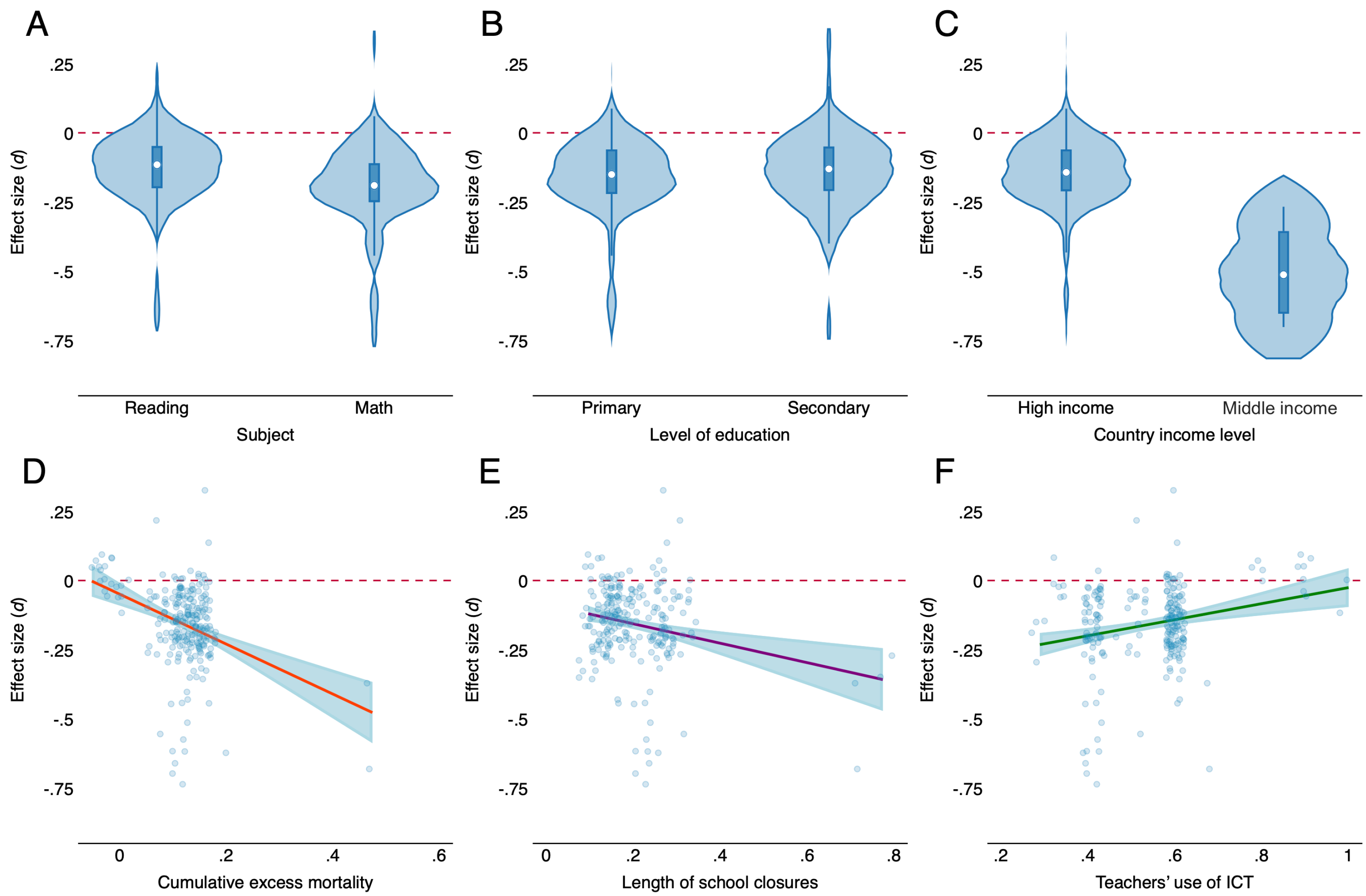


Figure 6: Variation in estimates of COVID-19 learning deficits (n=235) across individual- and country-level characteristics: (*A*) Learning subject (reading vs. math), (*B*) Level of education (primary vs. secondary), (*C*) Country income level (high vs. middle), (*D*) Excess mortality, (*E*) Length of school closures, (*F*) Teachers' use of information and communications technology (ICT). Panels (*A*), (*B*), and (*C*) each show the distribution of COVID-19 learning deficit estimates, with the box marking the interquartile range and the white circle denoting the median. Panels (*D*), (*E*), and (*F*) each show fitted regression lines with 95% confidence intervals. Points have been jittered to aid legibility.

larly when it is implemented without adequate preparation, teacher training and adequate digital learning technology in children's homes.

To examine the relationship between school closures and learning deficits, we collated data from the UNESCO school closures database[61] and the US School Closure and Distance Learning Database.[46] Fig. 6*E* plots all estimates of COVID-19 learning deficits against the cumulative length of school closures up until the date in the pandemic on which a given measure of learning deficit was taken. It suggests that estimated learning deficits are larger in contexts where face-to-face teaching was suspended for longer stretches of time.

Teacher use of ICT is associated with smaller learning deficits. The quality of digital learning infrastructure, as well as children's and teacher's ability to use it, is likely to influence the effectiveness of remote learning.[62] As such, it is a key concern in ongoing policy debates over how to recover lost and foregone learning, and how to limit further learning deficits in the future.

We examine the relationship between learning deficits and teachers' use of information and communications technology (ICT) in Fig. 6*F*. Information on ICT use in school was collected in the 2018 OECD Teaching and Learning International Survey (TALIS) and measures country-level variation in teachers' use of ICT prior to the pandemic (see *Methods*). Fig. 6*F* supports the expectation that teachers' experience with using ICT is helpful in limiting learning deficits. Learning deficits are smaller or absent in countries, such as Denmark,[6] where teachers report frequent ICT use.

Discussion

Two years into the COVID-19 pandemic, there is still insufficient knowledge about its consequences for the learning progress of school-age children. This paper makes two main contributions to better understand the state of the evidence. First, it systematically reviews the existing literature on the effect of the pandemic on learning among school-age children and appraises its geographic reach and quality. Second, it harmonizes, synthesizes and meta-analyzes the existing evidence in order to examine the extent to which the pandemic has affected learning, and how this varies across different groups of students and country contexts.

We identify a sizable and growing body of evidence on the effects of the pandemic on learning. However, existing studies primarily focus on high-income countries, while there is a dearth of evidence from low- and middle-income countries. This is particularly concerning because the small number of existing studies from middle-income countries suggest that learning deficits have been particularly severe in these countries. Learning deficits are likely to be even larger in low-income countries, considering that they already faced a learning crisis before the pandemic, generally implemented longer school closures, and were under-resourced and ill-equipped to facilitate remote learning. It is critical that this evidence gap on low- and middle-income countries is addressed swiftly, and that the infrastructure to collect and share data on educational performance in middle- and low-income countries is strengthened. Collecting and making available this data is a key prerequisite for fully understanding the effect of the pandemic on learning and related outcomes.[63]

About half of the studies that we identify are rated as having a serious risk of bias. Future studies should minimize risk of bias in estimating learning deficits by employing research designs that appropriately account for common sources of bias. These include a lack of accounting for secular time trends, non-representative samples, and imbalances between treatment and comparison groups. These potential sources of bias may lead existing studies to underestimate of learning deficits.

Our meta-analysis suggests that the COVID-19 pandemic has led to substantial learning deficits. The pooled effect size of $d = -0.17$, implies that students lost out on 0.17/0.4, or about 40%, of a school year's worth of learning. This confirms initial concerns that the pandemic would cause substantial harm to student learning.[35,64] But our results also show that fears of an accumulation of learning deficits as the pandemic continues have not materialized.[65,66] On average, learning deficits emerged early in the pandemic and have neither closed nor widened.

The persistence of learning deficits two years into the pandemic highlights the need for well-designed, well-resourced and decisive policy initiatives to recover learning deficits. Policy-makers, schools, and families will need to identify and realize opportunities to complement and expand on regular school-based learning. Promising avenues include the use of the often extensive summer holidays to offer summer schools and learning camps, extending school days and school weeks, and organizing and scaling up tutoring programs. Further potential lies in developing, improving, publicizing and providing access to learning apps, online learning platforms, or educational TV programs that are free at the point of use. Many countries have already begun investing significant resources to capitalize on these opportunities. If the momentum of these policy efforts is maintained and expanded, the disruptions to learning during the pandemic may prove to be a window of opportunity to improve and extend the education afforded to children.

Most studies that we identify find that learning deficits have been largest for children from disadvantaged socio-economic backgrounds. This holds across different time points during the pandemic, countries, grade levels, and learning subjects, and independently of how socio-economic background is measured. The pandemic has thus exacerbated educational inequalities that were already large before the pandemic.[67,68] Policy initiatives to compensate learning deficits need to prioritize support for children from low socio-economic backgrounds in order to allow them to recover the learning they lost during the pandemic.

Comparing estimates of learning deficits across subjects, we find that learning deficits tend to be larger in math than in reading. As noted above, this may be due to the fact that parents and children have been in a better position to compensate school-based learning in reading by reading at home. Accordingly, there are grounds for policy initiatives to prioritize the compensation of learning deficits in math and other science subjects.

Our analysis of macro-level moderators suggests that COVID-19 learning deficits are higher in countries that experienced a more severe course of the pandemic, as measured by cumulative excess mortality. This is likely to be due to the higher overall psychological and physical strain of the health crisis on children, parents, and teachers, which was intensified by restrictive measures to contain the spread of the virus. One important measure through which governments have sought to contain the spread of the virus has been the suspension of face-to-face teaching. Confirming warnings by experts early in the pandemic, our meta-analysis shows that longer school closures are associated with larger learning deficits. This finding highlights the importance of carefully weighing the costs to learning incurred by school closures against possible health benefits from containing the spread of infections. This is true particularly considering the mixed evidence on the effectiveness of school closures in containing the spread of COVID-19.[69,70]

While remote learning is likely to remain a poor substitute for face-to-face instruction, having an adequate infrastructure for digital learning is likely to be critical for limiting learning deficits in the event of school closures. We find that teachers' use of information and communication technology (ICT) prior to the onset of the pandemic moderates the effect of the pandemic on learning progress. Countries, such as Denmark,[6] in which ICT use was widespread before the pandemic saw little or no learning deficits during the pandemic. Governments should upgrade the infrastructure for digital learning, as this will not only guard against adverse consequences of future disruptions to schooling, but also provide a means to recovering existing learning deficits by complementing regular, school-based instruction.

Our analysis provides important evidence of how the COVID-19 learning deficit has varied between different groups of students and across country contexts. Given the limited reach of the existing evidence, we do not seek to identify the causal role of specific factors. A fruitful avenue for future research will be to use quasi-experimental designs to reveal how specific factors can account for individual and society-level variation in the extent of COVID-19 learning deficits. The considerable variation across subpopulations and country contexts highlights the need to better understand this variation, and identify mechanisms that can guide policy measures for limiting and counteracting learning deficits.

Methods

Eligibility criteria. We consider all types of primary research, including peer-reviewed publications, preprints, working papers, and reports for inclusion. To be eligible for inclusion, studies have to measure learning progress using test scores that can be standardized across studies using Cohen's *d*. Moreover, studies have to be in English, Danish, Dutch, French, German, Norwegian, Spanish or Swedish.

Search strategy and study identification. We identify eligible studies using the following steps. First, we developed a Boolean search string defining our population (school-aged children), exposure (the COVID-19 pandemic), and outcomes of interest (e.g., math and reading). The full search string can be found in the SI Appendix. We used this string to search the following academic databases: Coronavirus Research Database, Education Database, ERIC, International Bibliography of the Social Sciences (IBSS), Politics Collection (PAIS index, policy file index, political science database, and worldwide political science abstracts), Social Science Database, Sociology Collection (applied social science index [ASSIA] and abstracts, sociological abstracts, and sociology database), CINAHL, and Web of Science. Our initial search was conducted on April 27, 2021. Second, we hand-searched multiple preprint, working paper, and policy document repositories (SSRN, MPRA, IZA, NBER, OSF Preprints, PsyArXiv, SocArXiv, and EdArXiv) and relevant policy websites, including, but not limited to, the websites of the Organisation for Economic Co-operation and Development, the United Nations, the World Bank, and the Education Endowment Foundation. Third, we periodically posted our protocol via Twitter in order to crowdsource additional relevant studies not identified through the search. Last, to ensure that our analysis is comprehensive in terms of recent and relevant research, we conducted a final manual forward and backward citation search of all eligible studies identified by the above steps. This process was conducted on February 14, 2022.

Data extraction. From the studies that meet our inclusion criteria we extract all estimates of the effect of the pandemic on learning progress, separately for math and reading and for different school grades. We also extract the corresponding sample size, standard error, date(s) of measurement, author name(s), and country. Last, we record whether studies differentiate between children's socio-economic background, which measure is used to this end, and whether studies find an increase, decrease or no change in learning inequality. We contacted study authors, if any of the above information was missing in the study. Data extraction was performed by B.B. and validated independently by A.B.M., with discrepancies resolved through discussion and by conferring with P.E.

Measurement and standardization. We standardize all estimates of the effect of the pandemic on learning using Cohen's *d*, which expresses effect sizes in terms of standard deviations. Cohen's *d* is calculated as the difference in the mean learning gain in a given subject (math or reading) over two comparable periods before and after the onset of the pandemic, divided by the pooled standard deviation of learning progress in this subject:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$$

, where

$$s = \sqrt{\frac{(s_1^2 + s_2^2)}{2}}$$

Effect sizes expressed as β coefficients are converted to Cohen's d:

$$d = \frac{\beta}{se} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Subject. We use a binary indicator for whether the study outcome is math or reading. One study does not differentiate the outcome but includes a composite of math and reading scores.[15]

Level of education. We distinguish between primary and secondary education. We first consulted the original studies for this information. In cases where the level of education was undetermined, students' age was used in conjunction with information about education systems from external sources.[71]

Country income level. We follow the World Bank's classification of countries into four income groups: low, lower-middle, upper-middle, and high-income. Three countries in our sample place are in the upper-middle group: Brazil, Mexico and South Africa. Remaining countries are high-income.

Excess mortality. We use data on all-cause mortality from the World Mortality Dataset.[72] We first model monthly deaths in the years 2015–2019 as a function of 12 month indicators and a linear trend in year that are allowed to vary by country:

$$deaths_{ct} = \alpha_c + \sum_{m=1}^{12} \delta_{cm} month + \gamma_c year + \varepsilon_{ct}$$

We then extract predicted values from this model for the period 2020–2021, which we label $\widehat{deaths}_{ct}$ (c for country, t for time). Next, we calculate excess mortality from observed deaths:

$$excess_{ct} = \frac{deaths_{ct}}{\widehat{deaths}_{ct}} - 1$$

The variable we enter into our analysis is a cumulative average of estimated excess deaths from March 2020 until time t when learning was measured.

Length of school closures. We operationalize school closures as the share of time (starting in March 2020) that schools had been closed until learning was measured. Our primary source is the UNESCO school closures database.[61] We complement this information with data from the US School Closure and Distance Learning Database[46] that allows us so extract state-level averages for studies that sample a single US state.

Teachers' use of ICT. We use survey data on the percentage of teachers in a country who report regularly using information and communications technology (ICT) in teaching. This information is from the 2018 round of OECD's Teaching and Learning International Survey (TALIS).[73] The variable reflects the share of teachers who answer the question "I let students use ICT for projects or class work" with "frequently" or "always".

Data synthesis. We synthesize our data using three synthesis techniques. First, we use random-effects models to generate a forest plot, based on all available estimates of the effect of the pandemic on learning. We pool estimates using inverse variance weights to calculate an overall effect size (see Fig. 3).[74] Second, we code all estimates of the effect of the pandemic on educational inequality between children from different socio-economic backgrounds, according to whether they indicate a positive, negative, or no effect. We visualize the resulting distribution using a harvest plot (see Fig. 5).[75] Third, given that the limited country- and overtime variation precludes multi-variate or causal analyses, we examine the bivariate association between COVID-19 learning deficits and a range of individual- and country-level covariates, using a series of violin and scatter plots (see Fig. 6).

Pre-registration and data availability. We prospectively registered a protocol of our systematic review and meta-analysis in the International Prospective Register of Systematic Reviews (CRD42021249944). We will share all information and data needed to replicate our findings on the Open Science Framework (`https://osf.io/`) following journal publication.

Acknowledgements. B.B. gratefully acknowledges funding from the Observatoire Sociologique du Changement (OSC), Sciences Po. A.B.M. gratefully acknowledges funding from the Carlsberg Foundation. P.E. gratefully acknowledges funding from the Leverhulme Centre for Demographic Science, The Leverhulme Trust, and the Swedish council for working life and social research (grant no. 2016-07099).

References

[1] United Nations. *The impact of COVID-19 on children* (UN Policy Briefs, 2020).

[2] Gore, J., Fray, L., Miller, A., Harris, J. & Taggart, W. The impact of COVID-19 on student learning in New South Wales primary schools: an empirical study. *The Australian Educational Researcher* **48**, 605–637 (2021).

[3] Gambi, L. & De Witte, K. The resiliency of school outcomes after the COVID-19 pandemic: Standardised test scores and inequality one year after long term school closures. *FEB Research Report Department of Economics* (2021).

[4] Maldonado, J. E. & De Witte, K. The effect of school closures on standardised student test outcomes. *British Educational Research Journal* **48**, 49–94 (2021).

[5] Lichand, G., Dória, C. A., Neto, O. L. & Cossi, J. The impacts of remote learning in secondary education: Evidence from Brazil during the pandemic. Inter-American Development Bank (2021).

[6] Birkelund, J. F. & Karlson, K. B. No evidence of a major learning slide 14 months into the COVID-19 pandemic in Denmark. SocArXiv (2021).

[7] Depping, D., Lücken, M., Musekamp, F. & Thonke, F. Kompetenzstände Hamburger schüler*innen vor und während der corona-pandemie. In Fickermann, D. & Edelstein, B. (eds.) *Schule während der Corona-Pandemie. Neue Ergebnisse und Überblick über ein dynamisches Forschungsfeld*, 51–79 (Münster & New York: Waxmann, 2021).

[8] Schult, J. & Lindner, M. A. Did students learn less during the COVID-19 pandemic?: reading and mathematics competencies before and after the first pandemic wave. PsyArXiv (2021).

[9] Contini, D., Di Tommaso, L., Muratori, M. C., Piazzalunga, D. & Schiavon, L. The COVID-19 pandemic and school closure: Learning loss in mathematics in primary education. *IZA Discussion Papers* (2021).

[10] Hevia, F. J., Vergara-Lope, S., Velásquez-Durán, A. & Calderón, D. Estimation of the fundamental learning loss and learning poverty related to COVID-19 pandemic in Mexico. *International Journal of Educational Development* **88**, 102515 (2022).

[11] Engzell, P., Frey, A. & Verhagen, M. D. Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences* **118** (2021).

[12] Haelermans, C. *et al.* A full year COVID-19 crisis with interrupted learning and two school closures: The effects on learning growth and inequality in primary education. Maastricht University, Research Centre for Education and the Labour Market (ROA) (2021).

[13] Schuurman, T. M., Henrichs, L. F., Schuurman, N. K., Polderdijk, S. & Hornstra, L. Learning loss in vulnerable student populations after the first COVID-19 school closure in the Netherlands. *Scandinavian Journal of Educational Research* 1–18 (2021).

[14] Ardington, C., Wills, G. & Kotze, J. COVID-19 learning losses: Early grade reading in South Africa. *International Journal of Educational Development* **86**, 102480 (2021).

[15] Tomasik, M. J., Helbling, L. A. & Moser, U. Educational gains of in-person vs. distance learning in primary and secondary schools: A natural experiment during the COVID-19 pandemic school closures in Switzerland. *International Journal of Psychology* **56**, 566–576 (2021).

[16] Blainey, K., Hiorns, C. & Hannay, T. The impact of lockdown on children's education: A nationwide analysis (2020).

[17] Blainey, K. & Hannay, T. The impact of school closures on autumn 2020 attainment. RS Assessment from Hodder Education (2021).

[18] Blainey, K. & Hannay, T. The impact of school closures on spring 2021 attainment. RS Assessment from Hodder Education (2021).

[19] Blainey, K. & Hannay, T. The effects of educational disruption on primary school attainment in summer 2021. RS Assessment from Hodder Education (2021).

[20] Renaissance Learning, E. P. Understanding progress in the 2020/21 academic year: Complete findings from the autumn term. London: Department for Education (2021).

[21] Renaissance Learning, E. P. Understanding progress in the 2020/21 academic year: Initial findings from the spring term. London: Department for Education (2021).

[22] GL Assessment. Impact of COVID-19 on attainment: initial analysis. Brentford: GL Assessment (2021).

[23] Rose, S. *et al.* Impact of school closures and subsequent support strategies on attainment and socio-emotional wellbeing in key stage 1: Interim paper 1. London: National Foundation for Educational Research (NFER), Education Endowment Foundation (EEF).

[24] Rose, S. *et al.* Impact of school closures and subsequent support strategies on attainment and socio-emotional wellbeing in key stage 1: Interim paper 2. London: National Foundation for Educational Research (NFER), Education Endowment Foundation (EEF).

[25] Domingue, B. W., Hough, H. J., Lang, D. & Yeatman, J. Changing patterns of growth in oral reading fluency during the COVID-19 pandemic. PACE Working Paper, Policy Analysis for California Education (2021).

[26] Domingue, B. *et al.* The effect of COVID on oral reading fluency during the 2020-2021 academic year. EdArXiv (2021).

[27] Kogan, V. & Lavertu, S. The COVID-19 pandemic and student achievement on Ohio's third-grade English language arts assessment. Ohio State University (2021).

[28] Kogan, V. & Lavertu, S. How the COVID-19 pandemic affected student learning in Ohio: Analysis of spring 2021 Ohio state tests. Ohio State University (2021).

[29] Kozakowski, W., Gill, B., Lavallee, P., Burnett, A. & Ladinsky, J. Changes in academic achievement in Pittsburgh public schools during remote instruction in the COVID-19 pandemic (2020).

[30] Kuhfeld, M., Tarasawa, B., Johnson, A., Ruzek, E. & Lewis, K. Learning during COVID-19: Initial findings on students' reading and math achievement and growth. NWEA Brief (2020).

[31] Lewis, K., Kuhfeld, M., Ruzek, E. & McEachin, A. Learning during COVID-19: Reading and math achievement in the 2020-21 school year. NWEA Brief (2021).

[32] Lewis, K. & Kuhfeld, M. Learning during COVID-19: An update on student achievement and growth at the start of the 2021-22 school year. NWEA Brief (2021).

[33] Pier, L., Christian, M., Tymeson, H. & Meyer, R. H. COVID-19 impacts on student learning: Evidence from interim assessments in California. PACE Working Paper, Policy Analysis for California Education (2021).

[34] Sterne, J. A. *et al.* ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* **355** (2016).

[35] Azevedo, J. P., Hasan, A., Goldemberg, D., Iqbal, S. A. & Geven, K. Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates. *World Bank Policy Research Working Paper* (2020).

[36] Bloom, H. S., Hill, C. J., Black, A. R. & Lipsey, M. W. Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness* **1**, 289–328 (2008).

[37] Hill, C. J., Bloom, H. S., Black, A. R. & Lipsey, M. W. Empirical benchmarks for interpreting effect sizes in research. *Child development perspectives* **2**, 172–177 (2008).

[38] Belot, M. & Webbink, D. Do teacher strikes harm educational attainment of students? *Labour* **24**, 391–406 (2010).

[39] Jaume, D. & Willén, A. The long-run effects of teacher strikes: Evidence from Argentina. *Journal of Labor Economics* **37**, 1097–1139 (2019).

[40] Cygan-Rehm, K. Is additional schooling worthless? revising the zero returns to compulsory schooling in germany. *Revising the Zero Returns to Compulsory Schooling in Germany (August 23, 2018)* (2018).

[41] Ichino, A. & Winter-Ebmer, R. The long-run educational cost of world war ii. *Journal of Labor Economics* **22**, 57–87 (2004).

[42] Cooper, H., Nye, B., Charlton, K., Lindsay, J. & Greathouse, S. The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research* **66**, 227–268 (1996).

[43] Allington, R. L. *et al.* Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology* **31**, 411–427 (2010).

[44] Smith, W. C. Consequences of school closure on access to education: Lessons from the 2013–2016 ebola pandemic. *International Review of Education* **67**, 53–78 (2021).

[45] Andrabi, T., Daniels, B. & Das, J. Human capital accumulation and disasters: Evidence from the pakistan earthquake of 2005. *Journal of Human Resources* 0520–10887R1 (2021).

[46] Parolin, Z. & Lee, E. K. Large socio-economic, geographic and demographic disparities exist in exposure to school closures. *Nature human behaviour* **5**, 522–528 (2021).

[47] Goudeau, S., Sanrey, C., Stanczak, A., Manstead, A. & Darnon, C. Why lockdown and distance learning during the covid-19 pandemic are likely to increase the social class achievement gap. *Nature Human Behaviour* **5**, 1273–1281 (2021).

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

[49] van de Werfhorst, H. G. Inequality in learning is a major concern after school closures. *Proceedings of the National Academy of Sciences* **118** (2021).

[50] Alexander, K. L., Entwisle, D. R. & Olson, L. S. Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis* **23**, 171–191 (2001).

[51] Aucejo, E. M. & Romano, T. F. Assessing the effect of school days and absences on test score performance. *Economics of Education Review* **55**, 70–87 (2016).

[52] Gottfried, M. A. The detrimental effects of missing school: Evidence from urban siblings. *American Journal of Education* **117**, 147–182 (2011).

[53] Goodman, J. Flaking out: Student absences and snow days as disruptions of instructional time. National Bureau of Economic Research (2014).

[54] Angrist, N., Djankov, S., Goldberg, P. K. & Patrinos, H. A. Measuring human capital using global learning data. *Nature* **592**, 403–408 (2021).

[55] Torche, F. Educational mobility in the developing world 139–171 (2021).

[56] World Bank. *World development report 2018: Learning to realize education's promise* (The World Bank, 2018).

[57] United Nations. Policy brief: Education during COVID-19 and beyond (2020).

[58] UNESCO. One year into COVID-19 education disruption: Where do we stand? (2021).

[59] Azevedo, J. P., Hasan, A., Goldemberg, D., Geven, K. & Iqbal, S. A. Simulating the potential impacts of covid-19 school closures on schooling and learning outcomes: A set of global estimates. *The World Bank Research Observer* **36**, 1–40 (2021).

[60] Jackson, M. & Williams, J. L. Covid-19 mitigation policies and psychological distress in young adults. *SSM-Mental Health* **2**, 100027 (2022).

[61] UNESCO Institute of Statistics. Dashboards on the global monitoring of school closures caused by the COVID-19 pandemic. `https://covid19.uis.unesco.org/global-monitoring-school-closures-covid19/` (2022).

[62] Hossain, M. Unequal experience of covid-induced remote schooling in four developing countries. *International Journal of Educational Development* **85**, 102446 (2021).

[63] Conley, D. & Johnson, T. Opinion: Past is future for the era of covid-19 research in the social sciences. *Proceedings of the National Academy of Sciences* **118** (2021).

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

[65] Fuchs-Schündeln, N., Krueger, D., Ludwig, A. & Popova, I. The long-term distributional and welfare effects of COVID-19 school closures. National Bureau of Economic Research (2020).

[66] Kaffenberger, M. Modelling the long-run learning impact of the covid-19 learning shock: Actions to (more than) mitigate loss. *International Journal of Educational Development* **81**, 102326 (2021).

[67] Attewell, P. & Newman, K. S. *Growing gaps: Educational inequality around the world* (Oxford University Press, 2010).

[68] Betthäuser, B. A., Kaiser, C. & Trinh, N. A. Regional variation in inequality of educational opportunity across europe. *Socius* **7**, 23780231211019890 (2021).

[69] Ertem, Z. *et al.* The impact of school opening model on sars-cov-2 community incidence and mortality. *Nature Medicine* **27**, 2120–2126 (2021).

[70] Goldhaber, D. *et al.* To what extent does in-person schooling contribute to the spread of covid-19? evidence from michigan and washington. *Journal of Policy Analysis and Management* **41**, 318–349 (2022).

[71] Eurydice. Eurybase: the information database on education systems in europe (2021).

[72] Karlinsky, A. & Kobak, D. Tracking excess mortality across countries during the covid-19 pandemic with the world mortality dataset. *Elife* **10**, e69336 (2021).

[73] Ainley, J. & Carstens, R. Teaching and learning international survey (TALIS) 2018 conceptual framework. OECD (2018).

[74] Dekkers, O. M. *et al.* COSMOS-E: Guidance on conducting systematic reviews and meta-analyses of observational studies of etiology. *PLoS Medicine* **16**, e1002742 (2019).

[75] Ogilvie, D. *et al.* The harvest plot: A method for synthesising evidence about the differential effects of interventions. *BMC Medical Research Methodology* **8**, 1–7 (2008).

Supplementary Information for

A systematic review and meta-analysis of the impact of the COVID-19 pandemic on learning

Bastian A. Betthäuser, Anders M. Bach-Mortensen, and Per Engzell

Corresponding author: Bastian A. Betthäuser.
E-mail: bastian.betthaeuser@sciencespo.fr

This PDF file includes:

Supplementary text
Figs. S1 to S4
Table S1
References for SI reference citations

Contents

List of Figures

List of Tables

Supporting Information Text

1. Search

We identified eligible studies using four main steps. First, we developed the following Boolean search string defining our population (school-aged children), exposure (the COVID-19 pandemic), and outcomes of interest (e.g., math and reading). *Set 1: Corona* OR COVID* OR lockdown OR "lock down" OR pandemic; Set 2: student* OR pupil* OR child* OR education* OR school* OR learn* OR performance OR test* OR exam* OR study* OR reading OR math* OR spelling OR achievement* OR literacy OR inequalit* OR mobility OR attainment OR assessment* OR progress OR fluency OR growth OR arithmetic* OR ELA.*

We used this string to search the following academic databases: Coronavirus Research Database, Education Database, ERIC, International Bibliography of the Social Sciences (IBSS), Politics Collection (PAIS index, policy file index, political science database, and worldwide political science abstracts), Social Science Database, Sociology Collection (applied social science index [ASSIA] and abstracts, sociological abstracts, and sociology database), CINAHL, and Web of Science. Our initial search was conducted on April 27, 2021. Second, we hand-searched multiple preprint, working paper, and policy document repositories (SSRN, MPRA, IZA, NBER, OSF Preprints, PsyArXiv, SocArXiv, and EdArXiv) and relevant policy websites, including, but not limited to, the websites of the Organisation for Economic Co-operation and Development, the United Nations, the World Bank, and the Education Endowment Foundation. Third, we periodically posted our protocol via Twitter in order to crowdsource additional relevant studies not identified through the search. Last, to ensure that our analysis is comprehensive in terms of recent and relevant research, we conducted a final manual forward and backward citation search of all eligible studies identified by the above steps. This process was conducted on February 14, 2022.

2. Risk of bias assessment

As described in the main text, we assessed the quality of the evidence using an adapted version of the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool (1). More specifically, we analyzed the risk of bias of each estimate from confounding, sample selection, classification of treatments, missing data, the measurement of outcomes, and the selection of reported results. We then assigned each estimate an overall risk of bias rating (low, moderate, serious, or critical) based on the estimate and domain with the highest risk of bias. In line with ROBINS-I guidance, we excluded all studies rated 'critical' ($n = 2$) from our meta-analysis (1).

In the main text we show an overview of the distribution of all estimates of COVID-19 learning deficits according to their risk of bias rating. Fig. S1 is a traffic light plot showing the domain specific risk of bias ratings of estimates for each individual study. The plot displays the risk of bias ratings separately for each risk of bias domain, as well as the overall risk of bias rating of each study (rightmost column). Fig. S1 shows that common sources of potential bias were confounding, sample selection, and missing data. Table S1 provides further details on the main sources of potential bias for each study for which estimates of learning deficits were rated to be at serious risk of bias.

As explained in the main text, the likely consequence of the observed potential sources of bias would be an underestimation of COVID-19 learning deficits. Studies rated at risk of confounding typically compared only two time points, without accounting for longer time trends in learning progress. The main causes of selection bias were the use of convenience samples and/or insufficient consideration of self-selection by schools or students. Several studies found evidence of selection bias, often with students from a low socio-economic background or schools in deprived areas being underrepresented after—as opposed to before—the pandemic, but this was rarely adjusted for. Some studies also reported a higher amount of missing data post-pandemic, again generally without adjustment, and several studies did not report any information on missing data.

3. Assessment of publication bias

In the main manuscript we display a z-curve to test for the possibility of publication bias. Fig. S2 shows two additional graphs with the same purpose: a p-curve *(a)* and a funnel plot *(b)*. The p-curve in Fig. S2*(a)* plots the distribution of p-values of all extracted estimates (black line). If there was publication bias or p-hacking, one would expect a left-skewed distribution with most p-values being just below the p=0.05 threshold of statistical significance. We find a right-skewed distribution of p-values, indicating no evidence for publication bias or p-hacking.

The funnel plot in Fig. S2*(b)* shows the distribution of estimates (circles) by their estimated effect size (horizontal axis) and standard error (vertical axis). The mean effect size is indicated by the vertical red and dotted line. Funnel plots were developed for meta-analyses of randomized controlled trials (RCTs) that assume one underlying effect size. Following this assumption, one would expect estimates with small standard errors (high precision) to be near the mean effect size, and studies with larger standard errors (low precision) to be spread evenly on both sides of the mean, forming a roughly funnel-shaped distribution (demarcated by the diagonal red and dotted lines). Deviation from this funnel shape may indicate publication bias.

Given that the studies we examine in our meta-analysis are observational and not randomized, and given that one would expect strong heterogeneity in estimates of COVID-19 learning deficits across country contexts, the funnel plot is arguably less useful than the z-curve and the p-curve, shown above, for detecting publication bias in the studies we examine. Moreover, testing for funnel plot asymmetry is not recommended when performing a meta-analysis of similarly powered studies (2). This applies to the estimates in our meta-analysis, most of which are based on very large samples ($n > 10,000$ for 70% of estimates). Nonetheless, the absence of a clear pattern by which estimates deviate from the mean in Fig. S2*(b)* confirms that publication bias appears not to be a concern.

4. Robustness to country exclusions

Given the relatively low total number of estimates of learning deficits in our dataset, a potential concern is that the associations between learning deficits and different individual- and country-level characteristics shown in the main manuscript (Fig.6) are driven by individual countries. In Fig. S3 we therefore estimate bivariate Pearson correlations between learning deficits and each moderator variable, dropping one country in turn. The graphs in Fig. S3 show point estimates and 95% confidence intervals. Findings for country income level (middle vs. high), excess mortality, and teachers' use of information and communications technology (ICT) remain robust to dropping individual countries. For subject (math vs. reading), the association with the magnitude of learning deficits remains of a similar size but is no longer statistically significant when dropping estimates from the United States. This is likely due to the fact that the sample is halved when dropping estimates from the US (half of all estimates [n = 95] are from the US). Level of education (secondary vs. primary) does not show a statistically significant correlation with learning deficits in our main analysis, but does so—indicating lower learning deficits for secondary school students—once we exclude estimates from Mexico from the analysis. For length of school closures, the association is reduced in size and no longer statistically significant when excluding Mexico. In sum, our results remain largely robust to dropping individual countries from the analysis, but the absence of an association between learning deficits and grade level, and the association between learning deficits and the length of school closures appears to be partly driven by the inclusion of estimates from Mexico in the analysis.

5. School closure data

To examine the association between the length of school closures and learning deficits we use the UNESCO School Closures Database (UNESCO-SCD) (3) and the US School Closure and Distance Learning Database (US-SCDLD) (4). We complement UNESCO-SCD with US-SCDLD, as the US is (incorrectly) recorded as having no school closures in UNESCO-SCD. This is likely to be related to the fact that there was no nationwide school closure policy in the US. To extract the relevant information from US-SCDLD, we use the state-level dataset and define school closure as a 75% reduction or more in school attendance. Fig. S4 shows that the choice of data on US school closures does not drive our estimates. The regression slope retains a clear negative slope, regardless of whether we combine the UNESCO-SCD data with the US-SCDLD data (Fig. S4*(a)*) or treat US states as having implemented no school closures, as originally recorded in the UNESCO-SCD data (Fig. S4*(b)*).

Risk of bias domain

Study	Confounding	Selection of participants	Classification of treatments	Missing data	Measurement of outcomes	Selection of reported results	Overall risk of bias
Ardington et al. 2021	–	–	+	–	+	+	–
Bielinski et al. 2021	–	–	+	+	+	+	–
Birkelund et al. 2021	+	+	+	+	+	+	+
Blainey et al. 2020a	x	–	+	?	+	+	x
Blainey et al. 2021b	x	–	+	?	+	+	x
Blainey et al. 2021c	x	–	+	?	+	+	x
Blainey et al. 2021d	x	–	+	?	+	+	x
Clark et al. 2021	x	!	+	–	+	+	!
Contini et al. 2021	–	–	+	–	+	+	–
Department for Education 2021a	x	–	+	–	+	+	x
Department for Education 2021b	–	–	+	–	+	+	–
Depping et al. 2021	x	–	+	–	–	+	x
Domingue et al. 2021a	–	–	+	–	+	+	–
Domingue et al. 2021b	–	–	+	–	+	+	–
Engzell et al. 2021	+	+	+	+	+	+	+
Gambi and De Witte 2021	+	–	+	–	–	+	–
GL Assessment 2021	x	–	–	–	+	+	x
Gore et al. 2021	–	+	+	–	+	+	–
Haelermans et al. 2021	+	+	+	–	+	+	–
Hevia et al. 2022	x	x	+	–	+	+	x
Kogan and Lavertu 2021a	–	–	+	–	+	+	–
Kogan and Lavertu 2021b	+	–	+	–	+	+	–
Kozakowski et al. 2021	x	x	+	+	+	+	x
Kuhfeld et al. 2020	x	–	+	x	+	+	x
Lewis et al. 2021a	x	–	+	–	+	+	x
Lewis et al. 2021b	x	–	+	–	+	+	x
Lichand et al. 2021	+	+	+	+	+	+	+
Locke et al. 2021	–	–	+	–	+	+	–
Maldonado and De Witte 2020	–	+	+	–	+	+	–
Pier et al. 2021	–	–	+	–	+	+	–
Ramadhan and Suhendra 2021	!	!	+	–	+	–	!
Rose at al. 2021a	x	x	+	?	+	+	x
Rose et al. 2021b	x	x	+	?	+	+	x
Schult et al. 2021	–	–	–	–	+	+	–
Schuurman et al. 2021	x	–	+	x	+	+	x
Tomasik et al. 2020	–	+	+	–	+	+	–

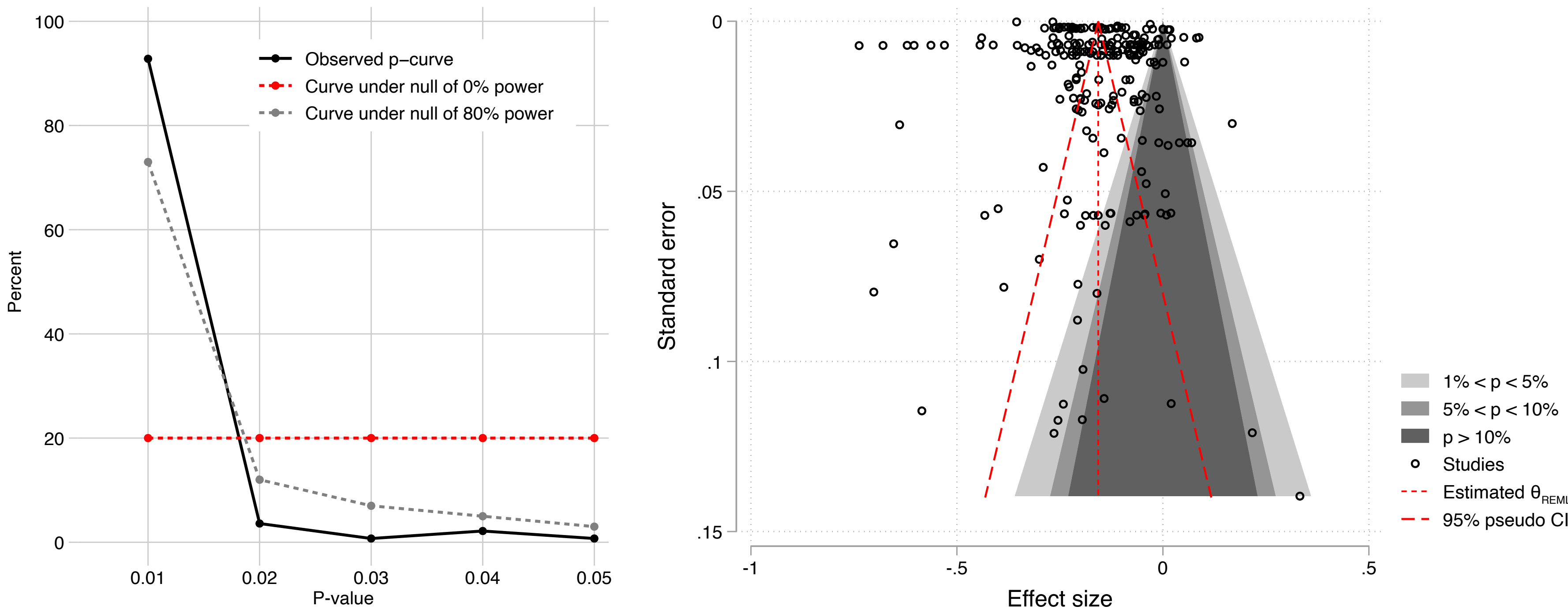
Risk of bias rating: + Low – Moderate ? No information x Serious ! Critical

Fig. S1. Risk of bias rating of estimates of COVID-19 learning deficits, by study, separately for each risk of bias domain (first 6 columns), and overall (final column). The risk of bias rating concerns the estimate(s) of COVID-19 learning deficits from each respective study, not to the study at large.

Table S1. Overview of potential sources of serious or critical risk of bias

Studies with estimates at serious risk of bias	Potential sources of bias
Blainey et al. 2020 (5) Blainey et al. 2021a-c (6–8)	The extracted estimates were not adjusted for secular time trends or other relevant confounding factors. The reports do not report sufficient information to enable assessment of the risk of bias related to the sample selection process or bias due to missing data.
Clark et al. 2021 (9), Ramadhan and Suhendra 2021 (10)	These studies constitute experimental studies with the primary aim to observe the effect of in-person vs. remote teaching on learning progress. While they contain information on learning deficits after the onset of the pandemic, the samples used were not designed to be representative of the society in question and are very small (containing students in 1 to 3 schools).
Department for Education 2021a-b (11, 12)	The report does not clearly state how relevant confounding factors were controlled for and does not include sufficient information to assess the risk of bias due to missing data.
Depping et al. 2021 (13)	The study reports high rates of missing data for the post-pandemic reading outcome. The measurement procedure differed for pre- and post-pandemic outcomes with pre-pandemic outcomes being assessed by external testing personnel and post-pandemic by teachers.
GL Assessment 2021 (14)	The study provides insufficient information to enable assessment of the risk of bias due to confounding and missing data.
Hevia et al. 2022 (15)	The extracted estimates are not adjusted for secular time trends and for all relevant confounding factors. The study does not include sufficient information to enable assessment of the risk of selection bias and bias due to missing data.
Kozakowski et al. 2021 (16)	The report does include enough information to enable assessment of the risk of confounding and selection bias.
Kuhfeld et al. 2020 (17), Lewis et al. 2021a-b (18, 19)	The extracted estimates are not adjusted for all relevant confounding factors and for the high levels of missing data in the post-pandemic sample.
Rose et al. 2021a-b (20, 21)	The extracted estimates are not adjusted for secular time trends and for all relevant confounding factors. The reports do not provide sufficient information to assess the risk of selection bias in their post-pandemic sample.
Schuurman et al. 2021 (22)	The extracted estimates are not adjusted for all relevant confounding factors and for the high rates of missing data in the post-pandemic outcomes.

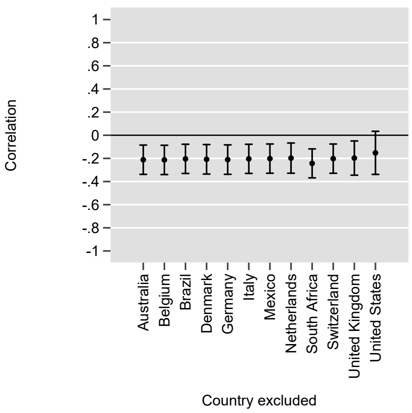
Note: Studies for which estimates of COVID-19 learning deficits are rated to be at serious risk of bias, with a description of potential sources of bias. The risk of bias rating relates to the estimate(s) of COVID-19 learning deficits from each respective study, not to the study at large. Full references are indicated by superscript and listed in the bibliography.



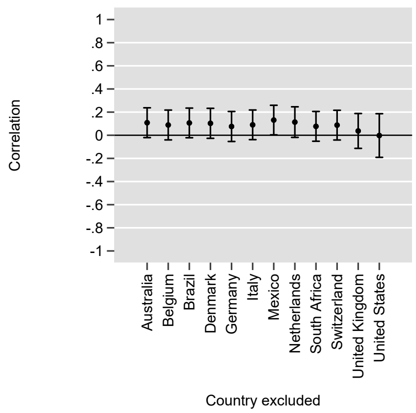
(a) *p*-curve

(b) Funnel plot

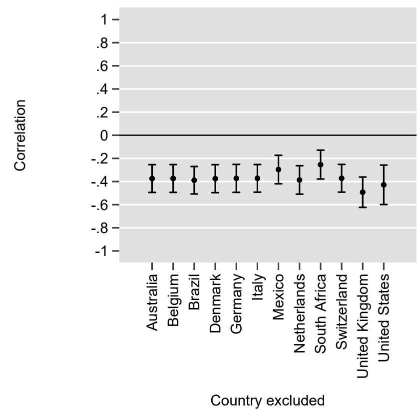
Fig. S2. Visual tests for publication bias. (*a*) *p*-curve: The black line shows the distribution of the *p*-values of all estimates to test for publication bias. If there was publication bias or *p*-hacking, one would expect a left-skewed distribution with most *p*-values being just below the *p*=0.05 threshold of statistical significance. The right skew of the distribution indicates no evidence for publication bias or *p*-hacking. (*b*) Funnel plot: The empty circles visualize the distribution of estimates by their estimated effect size (horizontal axis) and standard error (vertical axis). The mean effect size is indicated by the vertical red and dotted line. The absence of a clear pattern by which estimates deviate from the mean confirms that publication bias appears not to be a concern.



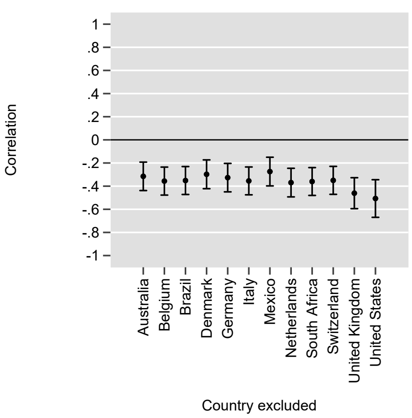
(a) Math vs. reading



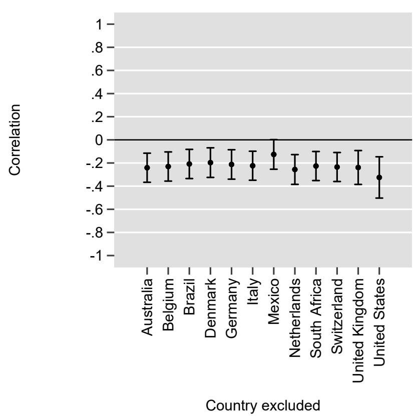
(b) Secondary vs. primary education



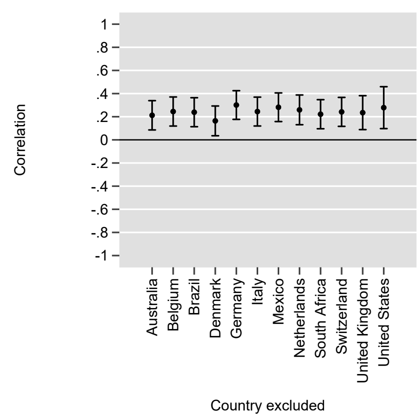
(c) Middle vs. high income country



(d) Excess mortality

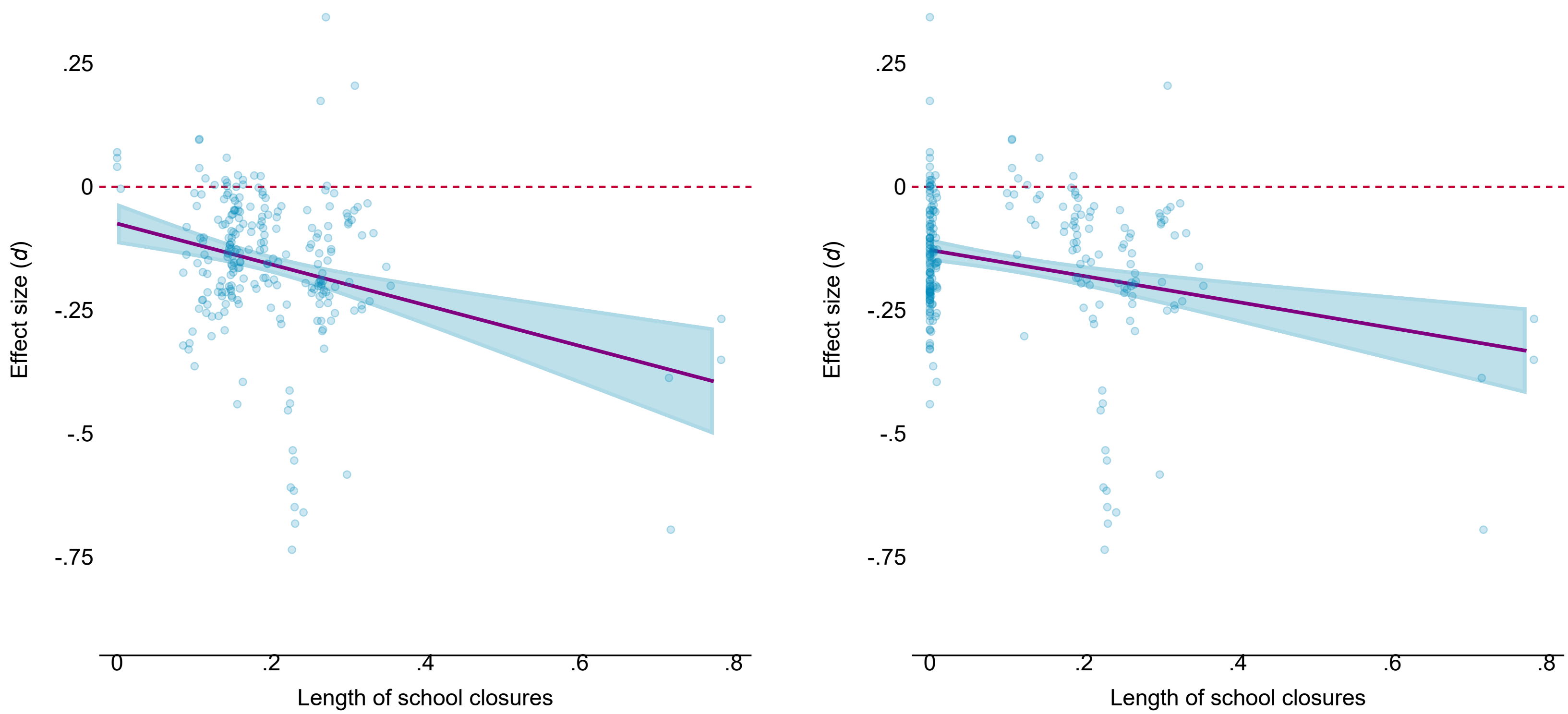


(e) Length of school closures



(f) Teachers' use of ICT

Fig. S3. Bivariate Pearson correlations between learning deficits and different moderator variables, dropping one country from the analysis in turn to test the robustness of the association in question: (*a*) Learning subject (math vs. reading), (*b*) Level of education (secondary vs. primary), (*c*) Country income level (middle vs. high), (*d*) Excess mortality, (*e*) Length of school closures, (*f*) Teachers' use of information and communications technology (ICT). The graphs show point estimates and 95% confidence intervals.



(a) UNESCO, US imputed

(b) UNESCO, US not imputed

Fig. S4. Variation in estimates of COVID-19 learning deficits (n=235) by length of school closure: (*A*) Using data from the UNESCO School Closures Database (UNESCO-SCD), combined with data on the length of school closures in the United States from the US School Closure and Distance Learning Database (US-SCDLD); (*B*) Using data from the UNESCO School Closures Database (UNESCO-SCD) only, which (incorrectly) indicates that there were no school closures in the United States.

References

1. JA Sterne, et al., ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* **355** (2016).
2. JA Sterne, et al., Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* **343**, d4002 (2011).
3. UNESCO Institute of Statistics, Dashboards on the global monitoring of school closures caused by the COVID-19 pandemic (https://covid19.uis.unesco.org/global-monitoring-school-closures-covid19/) (2022).
4. Z Parolin, EK Lee, Large socio-economic, geographic and demographic disparities exist in exposure to school closures. *Nat. human behaviour* **5**, 522–528 (2021).
5. K Blainey, C Hiorns, T Hannay, The impact of lockdown on children's education: A nationwide analysis. (2020).
6. K Blainey, T Hannay, The impact of school closures on autumn 2020 attainment (RS Assessment from Hodder Education) (2021).
7. K Blainey, T Hannay, The impact of school closures on spring 2021 attainment (RS Assessment from Hodder Education) (2021).
8. K Blainey, T Hannay, The effects of educational disruption on primary school attainment in summer 2021 (RS Assessment from Hodder Education) (2021).
9. AE Clark, H Nong, H Zhu, R Zhu, Compensating for academic loss: Online learning and student performance during the covid-19 pandemic. *China Econ. Rev*. **68**, 101629 (2021).
10. R Ramadhan, , et al., The effect of distance learning during covid-19 pandemics on the mathematical learning results in *Journal of Physics: Conference Series*. (IOP Publishing), Vol. 1806, p. 012096 (2021).
11. EP Renaissance Learning, Understanding progress in the 2020/21 academic year: Complete findings from the autumn term (London: Department for Education) (2021).
12. EP Renaissance Learning, Understanding progress in the 2020/21 academic year: Initial findings from the spring term (London: Department for Education) (2021).
13. D Depping, M Lücken, F Musekamp, F Thonke, Kompetenzstände Hamburger schüler*innen vor und während der corona-pandemie in *Schule während der Corona-Pandemie. Neue Ergebnisse und Überblick über ein dynamisches Forschungsfeld*, eds. D Fickermann, B Edelstein. (Münster & New York: Waxmann), pp. 51–79 (2021).
14. GL Assessment, Impact of COVID-19 on attainment: initial analysis (Brentford: GL Assessment) (2021).
15. FJ Hevia, S Vergara-Lope, A Velásquez-Durán, D Calderón, Estimation of the fundamental learning loss and learning poverty related to COVID-19 pandemic in Mexico. *Int. J. Educ. Dev*. **88**, 102515 (2022).
16. W Kozakowski, B Gill, P Lavallee, A Burnett, J Ladinsky, Changes in academic achievement in Pittsburgh public schools during remote instruction in the COVID-19 pandemic. (2020).
17. M Kuhfeld, B Tarasawa, A Johnson, E Ruzek, K Lewis, Learning during COVID-19: Initial findings on students' reading and math achievement and growth (NWEA Brief) (2020).
18. K Lewis, M Kuhfeld, E Ruzek, A McEachin, Learning during COVID-19: Reading and math achievement in the 2020-21 school year (NWEA Brief) (2021).
19. K Lewis, M Kuhfeld, Learning during COVID-19: An update on student achievement and growth at the start of the 2021-22 school year (NWEA Brief) (2021).
20. S Rose, et al., Impact of school closures and subsequent support strategies on attainment and socio-emotional wellbeing in key stage 1: Interim paper 1 (London: National Foundation for Educational Research (NFER), Education Endowment Foundation (EEF)) (year?).
21. S Rose, et al., Impact of school closures and subsequent support strategies on attainment and socio-emotional wellbeing in key stage 1: Interim paper 2 (London: National Foundation for Educational Research (NFER), Education Endowment Foundation (EEF)) (year?).
22. TM Schuurman, LF Henrichs, NK Schuurman, S Polderdijk, L Hornstra, Learning loss in vulnerable student populations after the first COVID-19 school closure in the Netherlands. *Scand. J. Educ. Res*., 1–18 (2021).