The Effect of School Closures on Standardized Test Scores: Evidence From a Zero-COVID Environment*

Christian Gillitzer
The University of Sydney

Nalini Prasad
UNSW Sydney

September, 2023

Abstract

Pandemic school closures were widespread but occurred together with COVID-19 health effects. Australia's successful COVID-19 elimination policies provide a unique setting to study the effect of school closures on learning loss absent significant health effects. We exploit variation in the duration of school closures across Australian regions of 9-157 school days and student-level test score data from a national standardized test with high participation to estimate learning loss. Learning loss was substantially smaller than comparable estimates from the literature, including for disadvantaged socioeconomic groups.

Keywords: student test scores, COVID-19, pandemic, NAPLAN, Australia

*We would like to thank Kadir Atalay, Colin Cameron, Daniel Carr, Ash Craig, Gigi Foster, Richard Holden, Greg Kaplan, Gregor Pfeifer, Stefanie Schurer, Mathias Sinning, participants at the 2023 Econometric Society Australasian Meetings and seminar participants at the University of Sydney and the Australian National University for helpful comments. The project received ethics approval by the UNSW Sydney and institutional exemption from ethics review by the University of Sydney. The NAPLAN data used in this study are available to academics, researchers and higher degree students at Australian tertiary institutions and authorized employees of Australian governments by application to the Australian Curriculum and Reporting Authority. This paper also uses data from the Household Income and Labour Dynamics in Australia (HILDA) Survey. See the disclaimer in the Supplementary Appendix for the HILDA survey. Authors: Christian Gillitzer, The University of Sydney, Sydney, Australia. Email address: christian.gillitzer@sydney.edu.au. Nalini Prasad, UNSW Sydney, Sydney, Australia. Email: nalini.prasad@unsw.edu.au.

1 Introduction

A narrative has developed that lost in person schooling during the pandemic caused substantial overall learning losses (Betthäuser, Bach-Mortensen and Engzell, 2023). However, school closures occurred together with high COVID-19 cases in most countries, making it difficult to isolate effects of school closures from the health effects of the pandemic. Australia provides a setting of international significance because lengthy school closures occurred when COVID-19 cases numbers were low, as part of Australia's successful COVID-19 elimination policy. In contrast to the existing literature, we find that learning losses caused by school closures were small.

Countries that pursued COVID-19 elimination policies included China, New Zealand, Singapore, South Korea, Taiwan and Japan. To the best of our knowledge, this paper provides the first evidence on learning loss from school closures in a zero-COVID country using results from a compulsory nationwide test. Australia closed its borders at the start of the pandemic and successfully implemented lockdowns and school closures to maintain zero COVID-19 until late 2021. Unlike in many other countries, lockdowns and school closures were implemented preemptively to stop the spread of COVID-19 cases rather than as a last-resort policy measure (Schurer et al., 2022). School closures in response to COVID-19 outbreaks ceased only after the adult population had reached 70 percent full COVID-19 vaccination (Fair Work Commission, 2022). Health effects of the pandemic were mild by international comparison (Mathieu et al., 2021). School closures occurred together with closure of non-essential workplaces and stringent restrictions on mobility outside the home. This restricted opportunities for non-participation in online learning and meant parents were often at home to supervise their child's learning.

It is challenging to isolate the effect of COVID-19 related school closures on student achievement because: (i) school closures typically affected all students within a jurisdiction providing no contemporaneous control group and (ii) student participation in tests fell noticeably during the pandemic (Werner and Woessmann, 2021). We address these issues by exploiting regional variation in COVID-19 outbreaks in Australia which caused exogenous contemporaneous variation in the duration of school closures which ranged from 9 to 157 days. We measure student achievement using results from a common standardized test. Results of this test are a strong predictor of a student's college entrance exam test scores (Houng and Justman, 2014). Test participation is compulsory for students in government and non-government schools and remained high during the pandemic. This minimizes non-participation bias. Testing occurred in person and under normal test conditions. An additional advantage of the Australian setting is that generous federal income support payments maintained incomes and employment, while the zero-COVID-19 policy pursued by the government meant that COVID-19 cases were low while schools were closed (Schurer et al., 2022). This gives us greater confidence that the effects we identify are from school closures, rather than the economic or health effects of the pandemic affecting student achievement.

We use cross-sectional variation in the duration of school closures in a difference-in-difference framework to estimate the causal effect of school closures on achievement using student-level data. Figure 1 previews our results, it shows average standardized test scores prior to and during the pandemic years by grade and region (ranked from those regions that experienced the longest to shortest duration of school closures). The variation in test scores we observe post school closures is like that observed in pre-pandemic years. We do not find evidence of large learning losses in regions which had longer durations of school closures relative to regions with a shorter duration. Our baseline causal difference-in-difference estimate is learning loss of 0.03 standard deviations per 100 days of school closures. This is 2-3 times smaller than comparable estimates in the literature. We find little evidence that school closures led to an increase in the number of students failing to meet minimum standards. We also find little evidence of large learning losses across most socio-economic groups, with the exception of indigenous students and those from a non-English speaking background. Our results are robust to different econometric specifications and controlling for student ability and test participation.

Much of the evidence on learning loss comes from studies in Europe at the start of the pandemic (Engzell, Frey and Verhagen, 2021; Maldonado and De Witte, 2022; Birkelund and Karlson, 2022; Tomasik, Helbling and Moser, 2021). See Betthäuser, Bach-Mortensen and Engzell (2023) for a review. Most of these papers benchmark achievement growth during the pandemic against achievement growth prior to the pandemic. This research design assumes that learning trajectories of the pandemic cohort would have been the same as earlier cohorts absent school closures. But health and other effects of the pandemic are likely to have disrupted learning absent school closures. This implies these studies provide an upper on the effect of school closures on learning loss. None of the 42 high-quality studies surveyed in Betthäuser, Bach-Mortensen and Engzell (2023) can exploit contemporaneous variation in the duration of school closures.

The few studies exploiting cross-sectional variation are for the United States. Jack et al. (2021) use state- and district-level variation in schooling mode but analyze differences in pass rates because tests differed across states. Like this study, Goldhaber et al. (2022) analyze student-level test score data, but study performance on a non-mandated test. To the best of our knowledge, this is the first study outside the United States exploiting cross-sectional variation in the duration of school closures. Unlike in the United States, schools in Australia had no discretion to choose instructional mode.

We consider the evidence on mechanisms contributing to limited learning loss. First, we document that there was widespread access to reliable internet, electronic devices and study space to facilitate home learning. Second, students continued to attend school when attendance shifted online: attendance in online learning was at or above the levels recorded prior to the pandemic. This is likely because students had fewer opportunities for non-participation in learning and did not miss school due to contracting the virus. Third, household survey data show increased parental inputs. Workplace closures meant parents were at home. Parents of children experiencing longer durations of school closures were more likely to reduce hours of paid work than parents of children experiencing short durations of school closures. Fourth, we find mixed evidence of temporary learning losses that faded over time.

2 Background

Australia pursued a policy of COVID-19 elimination until an adult vaccination rate of at least 70 percent had been achieved in each state and territory. The federal government closed the international border at the start of the pandemic in March 2020. The limited exemptions to the closed border, primarily Australians returning from overseas, were required to self-isolate in a hotel quarantine facility. State and territory governments were responsible for administering hotel quarantine systems.

The power to suspend in-person learning rests with the states in Australia. Key to our study design are differences both across and within states in the amount of time schools were closed. Individual schools had no power to choose instructional mode. All states implemented lockdowns and closed schools at the start of the pandemic in late March or early April 2020, but the return to in person learning and subsequent school closures varied in response to COVID-19 case numbers. All states re-opened schools by early June 2020 in a zero-COVID environment. Subsequently, states independently implemented lockdowns and school closures in response to detected COVID-19 cases. These subsequent COVID-19 cases were primarily caused by lapses in the hotel quarantine system. Borders were closed between states during periods of lockdowns and school closures. There were also restrictions on travel between metropolitan and regional areas of each state. This created variation in the duration of school closures by metropolitan/non-metropolitan region of each state. We exploit variation across all 14 regions, comprising two regions in each state (metro/non-metro) plus two territories.

To look at the effect of school closures on student achievement we examine the test scores of the 2021 and 2022 cohort of test takers. The 2021 cohort were subject to any school closures that took place from the start of 2020 to 11 May 2021 when the 2021 standardized exams were conducted. The 2022 cohort were subject to any school closures that took place from the start of 2020 to 10 May 2022 when the 2022 standardized exams were conducted. There was substantial variation in the duration of school closures across regions (Table 1). Over 2020 to 2021, the mean duration of school closures experienced by students was 77 days. The amount of time schools were closed for ranged from 9 days for a student in South Australia to 157 days for a student located in metropolitan Melbourne for the 2022 cohort. Students in metropolitan Victoria, regional Victoria, metropolitan New South Wales and the Australian Capital Territory all lost more than 75 days of in-person schooling. But students in South Australia, Western Australia and the Northern Territory lost fewer than 30 days of in-person schooling.

All school closures occurred during broader regional lockdowns which imposed stay at home orders and closed non-essential workplaces. In contrast to the international experience, lockdowns were a preventative measure. All lockdowns started when new daily infections were less than 100 and often fewer than 15. The peak 7-day average in new infections was 88 per million people until December 2021 (Mathieu et al., 2020). By December 2021, the vaccination rate was above 70 percent of the total population (Mathieu et al., 2020, 2021). The majority of COVID-19 deaths were among those aged over 70 years, so the health burden for school aged children and their families was low (Schurer et al., 2022). School closures occurred in the absence of widespread financial difficulty for households (Schurer et al., 2022). The provision of a generous federal wage subsidies mitigated profit and unemployment losses. The government also increased income-support payments. Hence, our results are unlikely to also capture negative health outcomes or income falls, which Kogan and Lavertu (2021) have shown affect student test scores.

The vast majority of students had access to sufficient technological resources to undertake studying from home. At least 96 percent of students reported that they had access to reliable internet, electronic devices and space for study. There was little difference in access to these resources across states and socio-demographic groups (see Appendix B.1 for more details). Hence, we can rule out that our results are being driven by differences in technological resources.

3 Data

3.1 School closures

We collect data on the duration of school closures from state government press releases and newspaper articles. We define a school as being closed for in-person learning if the government asked students to learn from home or if the duration of school holidays was extended. The latter typically lasted no longer than five days and primarily occurred at the start of the pandemic to allow teachers to prepare for remote learning. During school closures, teachers assigned students lessons to complete and checked in on students using online platforms (Ziebell et al., 2020).

3.2 Test scores

We measure student achievement from National Assessment Program—Literacy and Numeracy (NAPLAN) tests. Students sit NAPLAN tests in Grades 3, 5, 7 and 9. NAPLAN has been conducted annually in the second week of May since 2008, except 2020 when it was canceled due to the pandemic. The 2021 cohort of test takers experienced school closures that ranged from 4 to 109 days. The 2022 cohort of test takers experienced 9 to 157 days of school closures (Table 1).

All schools in Australia receive funding from government. A condition of funding is that students have to sit the NAPLAN tests, and school-level average NAPLAN test scores have to be published. Students sit standardized tests in reading, writing, spelling and grammar and numeracy. The tests are calibrated to a constant level of difficulty. Within a given subject area, a particular score represents the same level of achievement over time. The psychometric and scaling methods used to produce NAPLAN scores are similar to that used by the Programme for International Student Assessment. The writing test is graded by a person. Questions for all other tests have a specific answer and are graded by a computer.

We have access to de-identified student-level NAPLAN test score data for each year from 2013-2019 and 2021-2022 from the Australian Curriculum and Reporting Authority (ACARA). The dataset contains the population of students in Grades 3, 5, 7 and 9. There are test scores for between 1.05 and 1.2 million students in each year. We standardize test scores by the grade-level national standard deviation of test scores prior to the pandemic (2013-2019). The regions of Australia had comparable levels of test performance prior to the pandemic (Tables S1 and S2).[1]

The dataset contains information on student-level demographics including age, gender, indigenous status, language background other than English, highest level of school and post-school education for each parent and parental occupation. We combine education and occupation categories across parents by taking the higher group of either parent. The dataset records the test mode (paper or online) for each student and we control for this in our regression analyses.[2] There is information on school characteristics including the state and remoteness area (metropolitan, inner regional or outer regional/remote) of the school, school sector and a random school identifier. The data contains the record of all students including those who did not sit or abandoned the test, which we use to calculate participation rates. There are minimum standards of achievement defined for each grade level. Students below minimum standard are considered at risk of being unable to progress (ACARA, 2021).

Participation rates for the NAPLAN test have averaged over 90 percent for each grade level, including for the cohort of test takers affected by school closures (Figure S1). Test participation remained high in 2021 and 2022, mitigating concerns regarding selection.[3] Students in both 2021 and 2022 sat the NAPLAN test in person under normal testing conditions. For the 2021 cohort, low COVID-19 cases numbers meant that students across all states had returned to the classroom by October 2020 and spent the vast majority of the 2021 schooling year prior to the 2021 NAPLAN test in the classroom. For the 2022 cohort, all students had returned to the classroom by October 2021. Hence our results are unlikely to be affected by changes in test conditions or students being unaccustomed to the classroom.

4 Empirical methodology

We use individual test score data from the 2013–2019 and 2021-2022 NAPLAN tests and information on the number of days a student's school was closed owing to the COVID-19 pandemic to quantify the effect of school closures on student achievement. Using a difference-in-difference (DiD) model we compare the test scores of students in regions where schools were closed for a significant period of time to the test scores

[1] The exception is the Northern Territory, which has a large share of indigenous students.

[2] A transition from paper-based to online adaptive testing began in 2018 (ACARA, 2021). Scores for paper-based and online tests have been equated by the Australian Curriculum, Assessment and Reporting Authority (ACARA, 2021).

[3] Participation did fall in 2021 and 2022 in the Northern Territory but student numbers are small at around 1 per cent of students nationally.

for students in regions where schools were closed for a shorter period. We take advantage of substantial variation in the length of school closures across regions and between cohorts within regions. An event study model is used to validate our difference-in-difference design. We look for evidence of heterogeneous effects by socio-demographic characteristics by interacting the treatment indicator with each characteristic sequentially.

Baseline specification———We estimate the mean effect of learning from home on test scores using the two-way fixed effects (TWFE) model

$$s_{i,j,t} = \delta_j + \theta_t + \beta D_{j,t} + \sum_k \gamma_k X_{i,k} + \varepsilon_{i,j,t} \tag{1}$$

where $s_{i,j,t}$ is the standardized score for student i in region j and year t; δ_j is a region fixed effect; θ_t is a year fixed effect; and $D_{j,t}$ is the number of days of school closures in region j in year t. The coefficient of interest is β, the causal effect on test scores of an additional day of school closures. The term $\sum_k \gamma_k X_{i,k}$ is the set of covariates, which include the student-level demographic variables described earlier in Section 3.2 and test mode. $\varepsilon_{i,j,t}$ is an error term. We estimate a separate regression for each grade level as well as a regression pooled across grade levels; the pooled regression includes grade-level dummy variables as additional controls. We cluster standard errors at the region level. Because there are few clusters, we report standard errors using the Wild Cluster Bootstrap.[4] We estimate the regression separately for each test $s \in \{$Reading, Writing, Spelling and Grammar, Numeracy$\}$ and the composite (average) score across all tests. 5 percent of students sat at least one but not all of the tests. For these students, we compute the composite score by taking an average over the available test scores. We include dummy variables in the set of controls for each test s missing in the calculation of the composite score, to allow for differences in mean scores across tests.

Our baseline TWFE model pools data from two cohorts affected by school closures (2021 and 2022 test takers). In Section 5.3, we implement a stacked DiD model and show that our results are robust to the concerns raised by Goodman-Bacon (2021) and De Chaisemartin and d'Haultfoeuille (2020) about the TWFE model with staggered treatment timing. Callaway, Goodman-Bacon and Sant'Anna (2021) point out an additional issue that treatment effects may be biased when the treatment variable is continuous and units select into treatment based on potential outcomes. We can rule this out in our setting because school closures were a

[4]We impose the null of no effect. We have confirmed that p-values are similar for the Wild Cluster Restricted (null imposed) and Wild Cluster Unrestricted (null not imposed) bootstrap procedures, as required if standard errors are valid with few treated clusters (MacKinnon and Webb, 2018).

function of COVID-19 cases, which were exogenous to potential learning outcomes.

Event study specification———We use an event study model to estimate placebo effects prior to the pandemic. Because we have cohorts that took tests at difference times (2021 and 2022), we use a stacked event study design. The stacked event study gives an average of treatment effects for separate experiments (Cengiz et al., 2019). Our first experiment $d = 2021$ includes the pre-pandemic data 2013-2019 and the 2021 test takers, but not 2022 test takers; our second experiment $d = 2022$ includes the pre-pandemic data 2013-2019 and the 2022 test takers, but not 2021 test takers. The regression specification is

$$s^d_{i,j,t} = \delta^d_j + \theta^d_t + \sum_{t \neq 2019} \sum \beta_t \left(D^d_j \times I_t \right) + \sum_d \sum_k \gamma^d_k X^d_{i,k} + \varepsilon^d_{i,j,t} \tag{2}$$

where experiment $d \in \{2021, 2022\}$, I_t is a dummy variable equal to one for year t, D^d_j is the number of days of school closures experienced by students in region j and testing year 2021 (for experiment $d = 2021$) or testing year 2022 (for experiment $d = 2022$). Effects are relative to the base year 2019. All other terms are the same as in Equation (1). Finding the estimated pre-treatment effects $\beta_{t \neq 2019}$ to be insignificantly different from zero provides evidence that there were no confounding pre-trends prior to the pandemic.

Heterogeneous effects specification———We estimate heterogeneous treatment effects by characteristic k by re-specifying Equation (1) as follows:

$$s_{i,j,t} = \delta^k_j + \theta^k_t + \sum_k \beta^k \left(D_{j,t} \times I\left(i = k\right)\right) + \varepsilon_{i,j,t} \tag{3}$$

where k is the characteristic heterogeneity of interest (e.g. male/female) and $I(i = k)$ is a dummy variable taking the value one if student i has characteristic k. All other terminology is the same as in Equation (1). We consider each characteristics one at a time, excluding other control variables to minimize the chance of null results due to colinearity between characteristics.

5 Results

5.1 Graphical evidence

Figure 1 plots the composite test score for the pre-pandemic years 2013-2019 and the pandemic years 2021-22, by region and grade level. Data are de-meaned by the mean score in each region over the period 2013-2019. The pre-pandemic data (shown with gray circle markers) indicate the typical range of variation in scores by region absent school closures. An equating procedure is used by ACARA to ensure test are of similar level of difficulty over time. Regions are sorted from those that experienced the longest duration of school closures to those that experienced the shortest. The number of days of school closures experienced by region and cohort is shown next to markers in 2021 and 2022.

Test scores in 2021 and 2022 mostly fell within the range of past scores by region. The main exception is Grade 5 students where scores in 2021 and 2022 tended to be above previous scores.[5] There is little graphical evidence of larger declines in test scores in regions or cohorts that experienced longer durations of school closures. For 2021 test takers, students in Victoria-metro experienced school closures of 92-109 days but had test scores within the range of regions outside Victoria, which experienced school closures of no more than 42 days. Similarly, for 2022 test takers, the regions most affected by school closures, Victoria, New South Wales and the Australian Capital Territory had scores either above or within the range of other regions. These graphical results show no evidence of large declines in test scores in regions most affected by school closures, which we confirm in our causal effects analysis below.

5.2 Causal effect of school closures on test scores

Baseline results———Figure 2 presents estimates of our baseline model (Equation 1), by grade. Results are expressed per 100 days of school closures, approximately half the school year. Pooling data across grade levels, the mean learning loss per 100 days of school closures is 0.03 standard deviations (95 percent confidence interval [-0.07,0.06]). Composite learning loss is similar by grade level, and the confidence internals include no learning loss at each grade level except Grade 9. With few exceptions, learning loss is small across all test components (Figure 2). The exceptions are writing for Grade 3 students and numeracy for Grade 5 and 9 students. But there is no consistent pattern by test across grade levels.

Event study———The difference-in-difference model requires outcomes in the pre-pandemic years to provide a valid counterfactual. We use an event study model to look for evidence of confounding movements in test scores prior to the pandemic. The event study augments the baseline difference-in-difference model with lags of the treatment variable. Because we have two treated cohorts, the 2021 and 2022 test takers, we report an average of the responses for each cohort. Figure 3 presents the estimates of our stacked event study model, Equation (2), relative to the base year 2019. We find little evidence of learning loss. Test scores

[5] There were large gains in the Northern Territory in 2021 and 2022, but this is a small region, for which we could expect scores to be more variable. This raises estimated learning losses because the Northern Territory has one of the shortest duration of school closures.

for the pandemic years, 2021 and 2022, were within the range of prior scores. Few effects are statistically significant.

Heterogeneous effects———Learning loss could have been substantially larger for particular socio-demographic groups than the mean. We investigate this possibility by estimating separate treatment effects for different socio-demographic groups. Figure 4 shows estimates of the heterogeneous effects model, Equation (3). A separate regression is estimated for each group of characteristics, including only those covariates and test mode as controls. Learning loss is small for most socio-demographic characteristics, and is statistically insignificant in all cases (Figure 4). However, the point estimates indicate substantial learning losses for indigenous students and for students from LBOTE backgrounds.

Minimum standards———Students at risk of meeting minimum national benchmarks could have been most negatively affected by learning from home restrictions (Jack et al., 2021). To look for evidence of learning loss at the bottom of the test score distribution, we re-estimate our baseline model (Equation 1) replacing the left-hand size variable by an indicator variable taking the value one if the student meets the minimum standard. The effect per 100 days of school closures on the probability of meeting minimum standards is estimated to decline by less than 1 percentage point in all but two cases (Figure S2).

5.3 Robustness

We consider robustness of our baseline results to different estimation methodologies and controls.

Stacked difference-in-difference model———As discussed in Section 4, our baseline model uses variation between the 2021 and 2022 cohorts to estimate the effect of school closures on test scores. This can be problematic if treatment effects are heterogeneous across cohorts (De Chaisemartin and d'Haultfoeuille, 2020; Callaway, Goodman-Bacon and Sant'Anna, 2021). To check the robustness of our results to this concern, we estimate a stacked DiD model that averages treatment effects for two separate experiments, analogous to the stacked event study design described in Section 4. As in the stacked event study model, the first experiment $d = 2021$ includes the pre-pandemic data 2013-2019 and the 2021 test takers, but not 2022 test takers; the second experiment $d = 2022$ includes the pre-pandemic data 2013-2019 and the 2022 test takers, but not 2021 test takers. The stacked DiD model is

$$s_{i,j,t}^{d} = \delta_{j}^{d} + \theta_{j}^{d} + \beta D_{j,t}^{d} + \sum_{d} \sum_{k} \gamma_{k}^{d} X_{i,k}^{d} + \varepsilon_{i,j,t}^{d} \qquad (4)$$

where $d \in \{2021, 2022\}$, $D^d_{j,t}$ is the number of days of school closures for students in region j, time t and experiment d. Unlike the baseline model, the stacked DiD model uses only pre-pandemic (untreated) observations as counterfactuals to estimate β in each experiment. The stacked DiD results are shown in Figure S3. There is a negligible difference between the baseline and stacked DiD results.

Controlling for participation———Participation remained high in 2021 and 2022. Nevertheless, there has been some variation in participation across regions and over time. If non-participation is random, our results are unaffected. However, if more able students are more likely to participate than less able students our results will be biased. We include mean participation by region, year and grade level as an addition control, following Jack et al. (2021). The results are very similar to our baseline results (Figure S4). Note that enrollment changes between school sectors are not a concern because all students are required to sit NAPLAN.

Test mode———A transition from paper-based to online adaptive testing occurred between 2018 and 2021. Prior to 2018 all test were on paper and in 2022 all tests were online. ACARA equated paper-based and online tests and we have controlled for test mode in our analysis. As a further robustness check, we restrict the sample in each year to the most common test mode: paper-based from 2013-2018 and online for 2019 and 2021-2022. Estimates are very similar to our baseline results (Figures 2 and S5).

Controlling for previous score———Our dataset includes students' scores on NAPLAN tests taken two years prior, except for 2022 because NAPLAN tests were canceled in 2020 due to the pandemic. We re-specify the dependent variable in our baseline DiD model (Equation 1) to be the change in NAPLAN score. Comparing the change in score across regions controls for any correlation between region-level cohort effects and school closures. The point estimates are very similar to our baseline results (Figures 2 and S6). However, the confidence intervals are larger because after excluding the 2022 data there is less variation in school closures across regions.

Learning trajectories methodology———Most studies are not able to exploit cross-sectional variation in the length of school closures to estimate learning loss because school closures typically affected all children, providing no contemporaneous control group. Following Engzell, Frey and Verhagen (2021), many studies compare the growth in test scores of the pandemic cohort to previous cohorts. The identification assumption is that growth in scores of pre-pandemic cohorts provide a counterfactual for growth in scores during the pandemic absent school closures. We replicate this methodology using our dataset, comparing the performance of the 2021 cohort in Victoria-metro region to previous Victoria-metro cohorts. We restrict the sample to the Victoria-metro region because we do not have panel test score data for 2022 test takers and the Victoria-metro region experienced the longest duration of school closures among 2021 test takers. Following Equation (1) in Engzell, Frey and Verhagen (2021), the estimated regression is

$$\Delta s_{i,t} = \alpha + \beta D_{2021} + \sum_k \gamma_k X_{i,k} + \varepsilon_{i,t} \tag{5}$$

where $\Delta s_{i,t}$ is the change in composite NAPLAN score for student in between t and $t-2$, D_{2021} is a dummy variable for 2021 (the cohort of test takers affected by school closures) and all other terms are as for Equation (1). The controls include a linear time trend, replicating Engzell, Frey and Verhagen (2021). The learning loss estimates using this methodology are similar to our baseline results for each grade level (Figures 2 and S7).

5.4 Discussion

We find no evidence of large learning losses associated with school closures. The variation in test scores we observe post school closures is similar to that observed prior to the pandemic. Our baseline point estimates of learning loss are 0.03 standard deviations per 100 days of school closures; we can reject learning loss of more than 0.1 standard deviations across all grade levels. To put these estimates in perspective, the average gain in standardized test scores over 100 days for a student in Grade 5, 7 and 9 are 0.30, 0.18 and 0.14 standard deviations, respectively. This implies that per 100 days of school closures that the number of days of lost learning for a Grade 5, 7 and 9 student were 10, 17 and 21 days respectively. By way of comparison, prior to the pandemic a grade 5 and 7 student was on average absent from school for 7 days and a grade 9 student for 10 days per 100 days.

We can also benchmark our estimates against learning losses from summer holidays. Estimates at the bottom of the range suggest that during summer holidays student achievement falls by 0.1 standard deviations per 100 days (Downey, von Hippel and Broh (2004); Kuhfeld (2019); von Hippel and Hamrock (2019)). This is around 2-3 times larger than what we observe.

Our estimated learning loss from school closures during the pandemic is towards the lower end of estimates documented in the meta-analysis of Betthäuser, Bach-Mortensen and Engzell (2023). Pooling estimates across 42 studies, Betthäuser, Bach-Mortensen and Engzell (2023), estimate a learning loss of 0.14 standard deviations per school closure, roughly a learning loss of a third of a year. This is greater than our largest estimate across any individual grade level or subject. The estimates of Goldhaber et al. (2022) for the United States imply learning loss of 0.11 standard deviations per 100 days of remote schooling.[6] Unlike Jack et al. (2021) we find no evidence of large declines in the share of students meeting minimum standards.

The duration of school closures in Australia was towards the median of the countries mentioned in the above studies and longer than all school closures in Europe (Figure S8). A key difference in settings is that Australia pursued a zero-COVID policy. Australia had the strictest restrictions on human movement, based on the Oxford Coronavirus Government Response Tracker's Stringency Index (Mathieu et al., 2021).[7] Stay-at-home restrictions were in place 82 per cent of the time schools were closed, compared to an average of 64 per cent for the other countries (Figure S9). Australia's COVID-19 cases numbers were low while schools were closed, daily COVID-19 cases averaged 18.6 cases per million in Australia, while the next lowest number of daily COVID-19 cases was 62.3 cases per million (Figure S9).

Our result that the variation in test scores we observe post school closures is similar to previous years is consistent with results from Gore et al. (2021). Using data from New South Wales, the largest state of Australia, Gore et al. (2021) find no significant difference in the learning gains between the 2019 and 2020 cohorts.

Evidence of the effect of school closures on student achievement under zero-COVID policies is limited. Studies from Japan and China find little evidence of long term learning loss (Asakawa and Ohtake, 2022; Clark et al., 2021; Su et al., 2023). Evidence from South Korea suggests that school closures can widen the achievement gap between urban and rural students (Shin, An and Oh, 2023). But the datasets used in these papers cover a handful of schools or regions (Asakawa and Ohtake, 2022; Shin, An and Oh, 2023; Clark et al., 2021) or measure achievement using parental assessments (Su et al., 2023).

6 Mechanism Exploration

Why don't we find large learning losses associated with school closures? There is evidence that lost instructional time owing to extreme weather (Marcotte and Hemelt, 2008; Hansen, 2011; Miller and Hui, 2022), teacher strikes (Jaume and Willén, 2019; Belot and Webbink, 2010) and shortening the school year (Fitz-

[6] Table 1 of Goldhaber et al. (2022) reports learning loss of -0.201 standard deviations for 100 percent remote schooling in 2020-21. Assuming 180 school days per year, this equates to learning loss of 0.11 standard deviations per 100 days of remote schooling.

[7] This index measures the strictness of policies to restrict human movement in terms of school, workplace and public transport closures, restrictions on public gatherings, stay-at-home requirements, public information campaigns and travel restrictions.

patrick, Grissmer and Hastedt, 2011; Hansen, 2011; Pischke, 2007) lowers student achievement. In these studies, learning stopped while schools were closed. In our case, students were expected to learn at home.

School closures in Australia occurred during broader regional lockdowns which imposed stay-at-home orders. As noted in Section 2, the vast majority of students had access to sufficient technological resources to study from home. We find evidence that the vast majority of students continued to participate in learning when it was shifted online. Schools kept daily records of student attendance in online learning. Students were marked present if they returned schoolwork, participated online or answered a roll call (ACARA, 2022*a*). Attendance in online learning was at or above typical levels for in-person schooling (Table S4).[8] High attendance rates were likely due to low COVID-19 cases, so students did not miss school due to contracting the virus and stay-at-home restrictions provided fewer opportunities for activities outside the home. Surveys indicate that students spent on average 4 hours a day on schooling while learning from home, which is only slightly less than the five hours per day students spend in the classroom during a regular year (Bower, Lai and Van Bergen 2021; Australian Institute of Family Studies 2021).

Stay-at-home orders and associated workplace closures were used more frequently in Australia compared to other countries (Mathieu et al., 2021). This meant that most parents worked from home and could supervise their child's learning. Models of human capital accumulation indicate that increased parental teaching effort can mitigate the negative effect of school closures (Fuchs-Schündeln et al., 2022). Surveys suggest that parents spent around 14 hours per week supervising learning (Bower, Lai and Van Bergen 2021). Parental involvement in learning also increased the longer the duration of school closures. In households where both parents worked, the probability that at least one parent reduced the amount of paid work they undertook increased by close to 10 percentage points when the duration of learning from home increased from its median value to the 90th percentile (see Appendix B.2).

Nevertheless, it could still be the case that learning from home is less effective than learning in a classroom. There could have been temporary learning losses associated with school closures which unwound with the resumption of in-person schooling. Evidence from New South Wales is mixed. "Check-in" tests, with a similar format and psychometric properties to NAPLAN, were conducted on average around 50 days after students had returned to the classroom and around 100 days before the NAPLAN exam. NSW Department of Education (2020) and NSW Department of Education (2021) compared the affected cohorts check-in scores

[8] Attendance in all regions fell below usual levels in Term 1 2022, during which all regions had in-person schooling. This is because Australia's first significant COVID-19 wave, albeit in a vaccinated population, occurred in early 2022.

to previous cohorts' NAPLAN results and found suggestive evidence of learning loss. However, Gore et al. (2021), using data from a different test administered in NSW, finds little evidence of learning loss.

7 Conclusion

Australia provides a unique setting of international significance to estimate the effect of pandemic school closures on standardized test scores. Australia successfully pursued a COVID-elimination strategy, with states independently implementing school closures as part of broader lockdowns to suppress community transmission of COVID-19. This provides contemporaneous variation in the duration of school closures that is plausibly exogenous to the school system. There were large differences in the durations of school closures across regions: students in metropolitan Victoria experienced 147-157 days of school closures between 2020 to 2021, compared with only 9 days for students in South Australia. We measured student achievement from a common compulsory test with high participation.

We find that school closures caused small and statistically insignificant leaning losses. The variation in test scores we observed between regions post-pandemic is similar to that observed in prior years. We find no evidence of large learning losses for most disadvantaged socio-economic groups. Our results demonstrate that learning losses from school closures are not necessarily large in a country pursuing zero-COVID policies. Potential mechanisms include parental substitution for teachers and catch-up in the classroom. Further work should seek to provide a deeper understanding of these mechanisms.

This paper does not imply that school closures have zero costs. It is possible that school closures lowered student achievement in domains not tested and had negative effects on students' social skills.

References

ACARA. 2021. "NAPLAN Achievement in Reading, Writing, Language Conventions and Numeracy: National Report for 2021." Australian Curriculum, Assessment and Reporting Authority.

ACARA. 2022*a*. "Standard responses to attendance frequently asked questions during covid-19 pandemic." Australian Curriculum and Reporting Authority. Available at: <www.acara.edu.au/docs/default-source/default-document-library/faqs—recording-attendance-during-covid-19-pandemic.pdf>.

ACARA. 2022*b*. "Student attendance." Australian Curriculum and Reporting Authority. Available at: <www.acara.edu.au/reporting/national-report-on-schooling-in-australia/student-attendance>.

Asakawa, Shinsuke, and Fumio Ohtake. 2022. "Impact of COVID-19 School Closures on the Cognitive and Non-cognitive Skills of Elementary School Students." Research Institute of Economy, Trade and Industry (RIETI).

Australian Institute of Family Studies. 2021. "The Longitudinal Study of Australian Children." Australian Institute of Family Studies.

Belot, Michèle, and Dinand Webbink. 2010. "Do teacher strikes harm educational attainment of students?" *Labour*, 24(4): 391–406.

Betthäuser, Bastian A, Anders M Bach-Mortensen, and Per Engzell. 2023. "A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic." *Nature Human Behaviour*, 1–11.

Birkelund, Jesper Fels, and Kristian Bernt Karlson. 2022. "No evidence of a major learning slide 14 months into the COVID-19 pandemic in Denmark." *European Societies*, 1–21.

Bower, M, J Lai, and P Van Bergen. 2021. "Parent perceptions of COVID19 Schooling from home." Macquarie School of Education.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with a continuous treatment." *arXiv preprint arXiv:2107.02637*.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The effect of minimum wages on low-wage jobs." *The Quarterly Journal of Economics*, 134(3): 1405–1454.

Clark, Andrew E, Huifu Nong, Hongjia Zhu, and Rong Zhu. 2021. "Compensating for academic loss: Online learning and student performance during the COVID-19 pandemic." *China Economic Review*, 68: 101629.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review*, 110(9): 2964–2996.

Downey, Douglas B., Paul T. von Hippel, and Beckett A. Broh. 2004. "Are Schools the Great Equalizer? Cognitive Inequality during the Summer Months and the School Year." *American Sociological Review*, 69(5): 613–635.

Engzell, Per, Arun Frey, and Mark D Verhagen. 2021. "Learning loss due to school closures during the COVID-19 pandemic." *Proceedings of the National Academy of Sciences*, 118(17): e2022376118.

Fair Work Commission. 2022. "Information note - Government responses to COVID-19 pandemic." Fair Work Commission. Available at: <www.fwc.gov.au/about-us/news-and-media/news/covid-19-information-notes-and-updates>.

Fitzpatrick, Maria D, David Grissmer, and Sarah Hastedt. 2011. "What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment." *Economics of Education Review*, 30(2): 269–279.

Fuchs-Schündeln, Nicola, Dirk Krueger, Alexander Ludwig, and Irina Popova. 2022. "The long-term distributional and welfare effects of Covid-19 school closures." *The Economic Journal*, 132(645): 1647–1683.

Goldhaber, Dan, Thomas J Kane, Andrew McEachin, Emily Morton, Tyler Patterson, and Douglas O Staiger. 2022. "The consequences of remote and hybrid instruction during the pandemic." National Bureau of Economic Research.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277.

Gore, Jennifer, Leanne Fray, Andrew Miller, Jess Harris, and Wendy Taggart. 2021. "The impact of COVID-19 on student learning in New South Wales primary schools: an empirical study." *The Australian Educational Researcher*, 48(4): 605–637.

Hansen, Benjamin. 2011. "School year length and student performance: Quasi-experimental evidence." *Available at SSRN 2269846*.

Houng, B, and M Justman. 2014. "NAPLAN Scores as Predictors of Access to Higher Education in Victoria." Melbourne Institute Working Paper 22/14.

Jack, Rebecca, Clare Halloran, James C Okun, and Emily Oster. 2021. "Pandemic schooling mode and student test scores: Evidence from US states." National Bureau of Economic Research.

Jaume, David, and Alexander Willén. 2019. "The long-run effects of teacher strikes: evidence from Argentina." *Journal of Labor Economics*, 37(4): 1097–1139.

Kogan, V, and S Lavertu. 2021. "The COVID-19 pandemic and student achiement on Ohio's Third Grade English Language Arts Assessment." Ohio State University Applied Report.

Kuhfeld, Megan. 2019. "Surprising new evidence on summer learning loss." *Phi Delta Kappan*, 101(1): 25–29.

MacKinnon, James G, and Matthew D Webb. 2018. "The wild bootstrap for few (treated) clusters." *The Econometrics Journal*, 21(2): 114–135.

Maldonado, Joana Elisa, and Kristof De Witte. 2022. "The effect of school closures on standardised student test outcomes." *British Educational Research Journal*, 48(1): 49–94.

Marcotte, Dave E, and Steven W Hemelt. 2008. "Unscheduled school closings and student performance." *Education Finance and Policy*, 3(3): 316–338.

Mathieu, Edouard, Hannah Ritchie, Esteban Ortiz-Ospina, Max Roser, Joe Hasell, Cameron Appel, Charlie Giattino, and Lucas Rodés-Guirao. 2021. "A global database of COVID-19 vaccinations." *Nature human behaviour*, 5(7): 947–953.

Mathieu, Edouard, Hannah Ritchie, Lucas Rodes-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina, and Max Roser. 2020. "Coronavirus Pandemic (COVID-19)." *Our World in Data*. https://ourworldindata.org/coronavirus.

Miller, Rebecca K, and Iris Hui. 2022. "Impact of short school closures (1–5 days) on overall academic performance of schools in California." *Scientific reports*, 12(1): 1–13.

NSW Department of Education. 2020. "Check-in assessments: Years 3, 5 and 9." NSW Department of Education.

NSW Department of Education. 2021. "2021 Term 2 and Term 4 Check-in assessments." NSW Department of Education.

Pischke, Jörn-Steffen. 2007. "The impact of length of the school year on student performance and earnings: Evidence from the German short school years." *The Economic Journal*, 117(523): 1216–1242.

Schurer, Stefanie, Kadir Atalay, Nick Glozier, Esperanza Vera-Toscano, and Mark Wooden. 2022. "Zero-COVID Policies: Melbourne's 112-Day Hard Lockdown Experiment Harmed Mostly Mothers." *medRxiv*.

Shin, Chol-Kyun, Youngeun An, and Soon-young Oh. 2023. "Reduced in-person learning in COVID-19 widens student achievement gaps in schools." *Asia Pacific Education Review*, 1–11.

Su, Y, X Anping, D Gong, and A Yan. 2023. "The Impacts of The Zero-Covid Policy on Academic Performance: Evidence from Primary and Secondary Students in China."

Tomasik, Martin J, Laura A Helbling, and Urs Moser. 2021. "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(4): 566–576.

UNESCO. 2022. "UNESCO map on school closures." UNESCO.

von Hippel, Paul T., and Caitlin Hamrock. 2019. "Do Test Score Gaps Grow Before, During, or Between the School Years? Measurement Artifacts and What We Can Know in Spite of Them." *Sociological Science*, 6(3): 43–80.

Werner, Katharina, and Ludger Woessmann. 2021. "The legacy of COVID-19 in education." IZA Institute of Labor Economics Discussion Paper 14796.

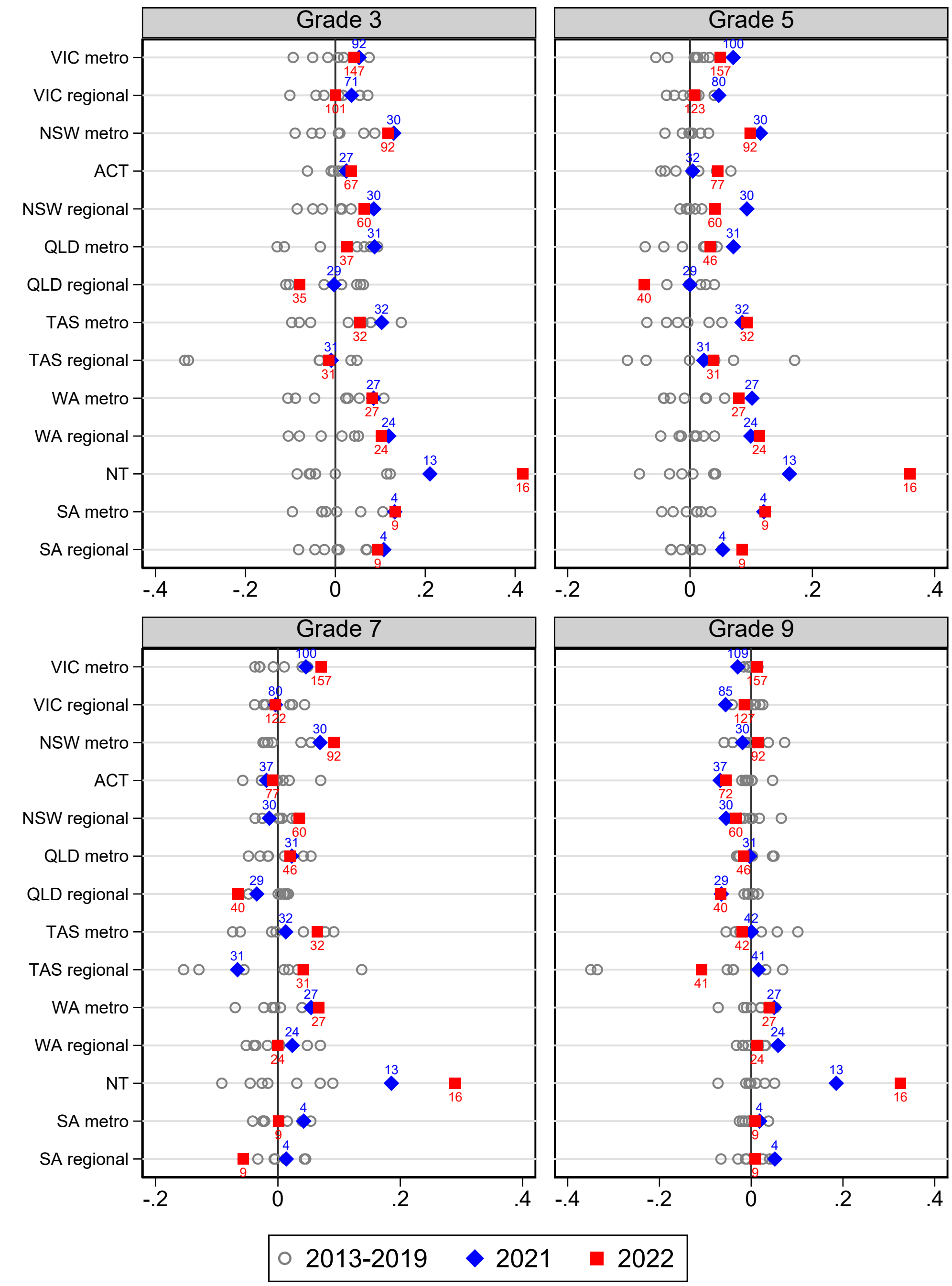
Ziebell, Natasha, Daniela Acquaro, Cath Pearn, and Wee Tiong Sean. 2020. "Examining the impact of COVID-19 Report Summary." Melbourne Graduate School of Education.

Table 1: Number of Days of School Closures: By Region, Cohort and Grade Level

	2021 Test Takers					2022 Test Takers				
	N	G3	G5	G7	G9	N	G3	G5	G7	G9
NSW metro	292,637	30	30	30	30	290,856	92	92	92	92
NSW regional	90,723	30	30	30	30	88,901	60	60	60	60
VIC metro	230,064	92	100	100	109	231,034	147	157	157	157
VIC regional	67,886	71	80	80	85	66,946	101	123	122	127
QLD metro	162,105	31	31	31	31	162,227	37	46	46	46
QLD regional	88,012	29	29	29	29	87,093	35	40	40	40
WA metro	105,748	27	27	27	27	106,016	27	27	27	27
WA regional	27,854	24	24	24	24	27,455	24	24	24	24
SA metro	58,829	4	4	4	4	59,604	9	9	9	9
SA regional	19,778	4	4	4	4	19,576	9	9	9	9
TAS metro	18,275	32	32	32	42	18,356	32	32	32	42
TAS regional	6,614	31	31	31	41	6,407	31	31	31	41
ACT	22,207	27	32	37	37	21,893	67	77	77	72
NT	11,018	13	13	13	13	10,582	16	16	16	16

Notes: The table shows the number of days schools were closed from January 2020 until the date of testing, by region and grade level. The 2021 cohort took NAPLAN tests in the period 11-21 May 2021 and the 2022 cohort took tests in the period 10-20 May 2022. N is the number of test takers by region and year, across all grades. Where there was variation within regions, the table reports a student-weighted average. There was some minor variation by school type (combined vs. primary/secondary) that is included in the analysis but not reported here.

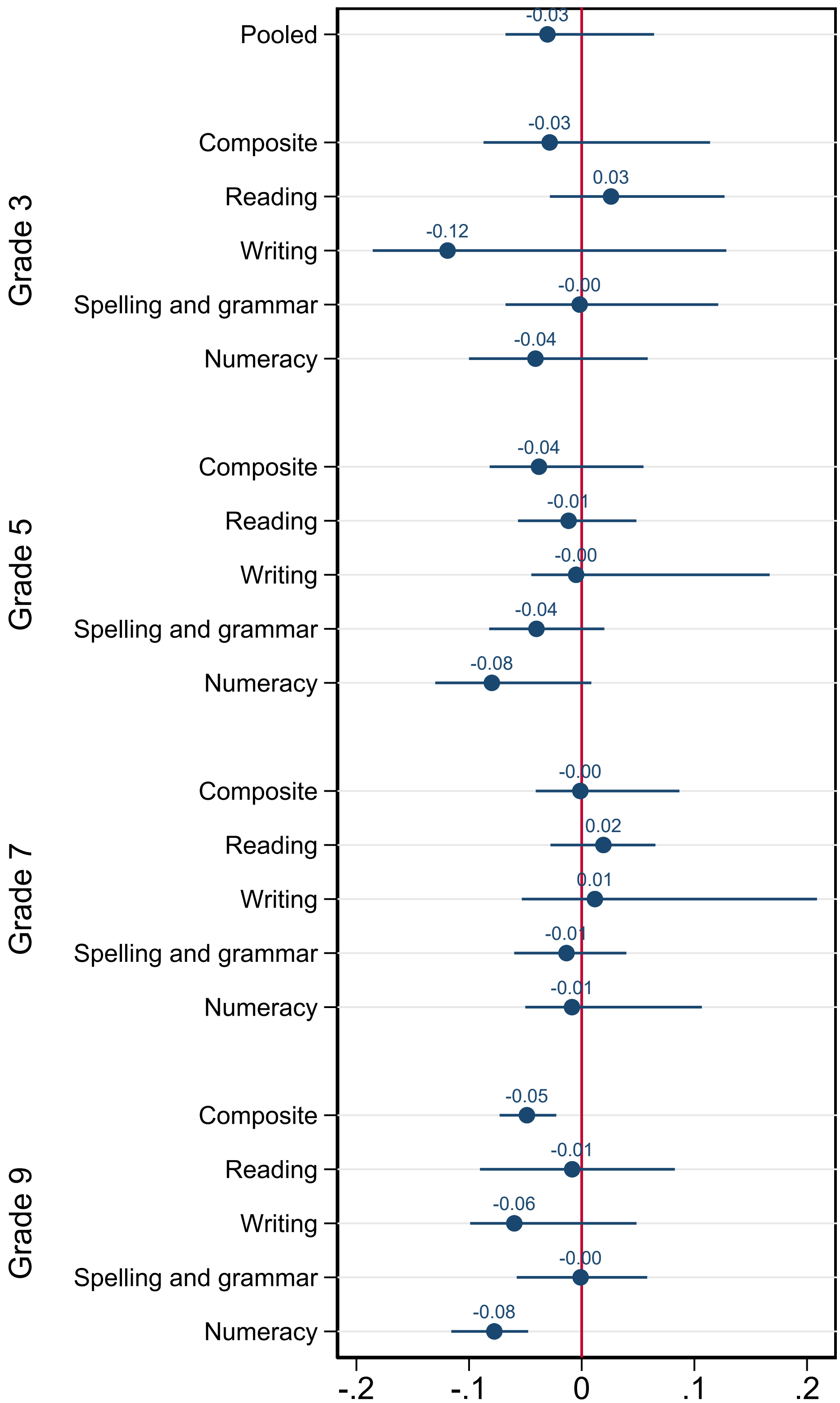
Figure 1: Mean NAPLAN Scores by Region and Year



Notes: Markers show mean NAPLAN scores by region and grade level. Years 2013-2019 are show by circle markers, 2021 by diamond markers and 2022 by square markers. The number of days of school closures is shown above markers for 2021 test takers and below markers for 2022 test takers. Scores are standardized by the grade-level national standard deviation over the period 2013-2019 and shown relative to each region's mean over the period 2013-2019.

Summary: Variation in mean scores was similar across regions with long and short durations of school closures.

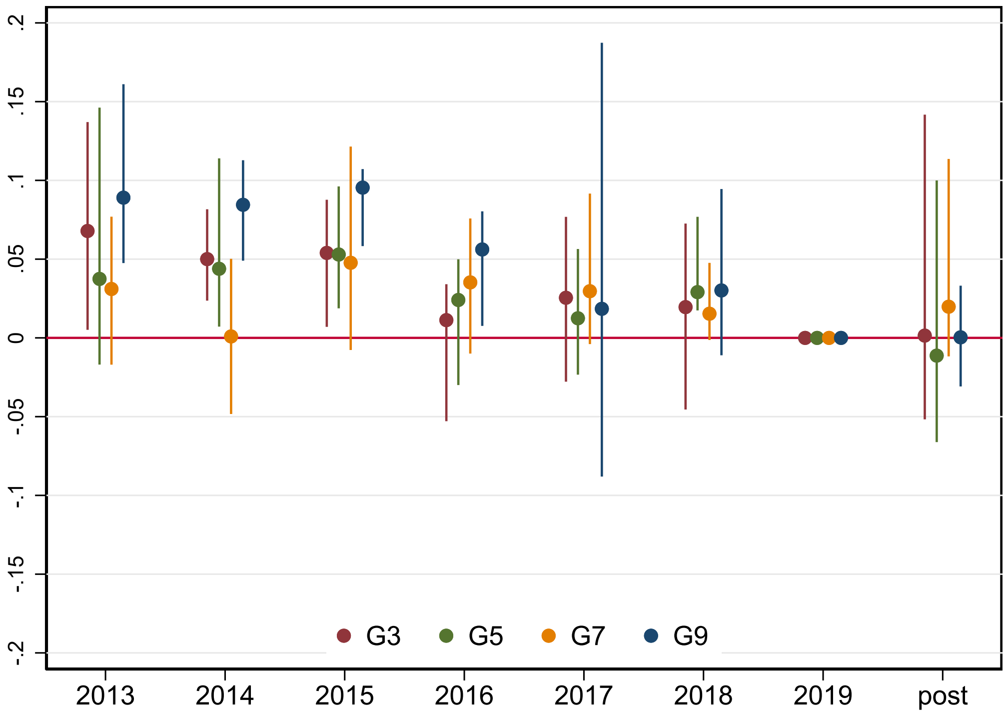
Figure 2: Learning Loss: Per 100 days School Closures



Notes: The figure shows estimated β coefficients for Equation (1) for the composite NAPLAN score and each test component. Pooled is the composite score for pooled grade levels. Bar range shows 95 percent confidence interval.

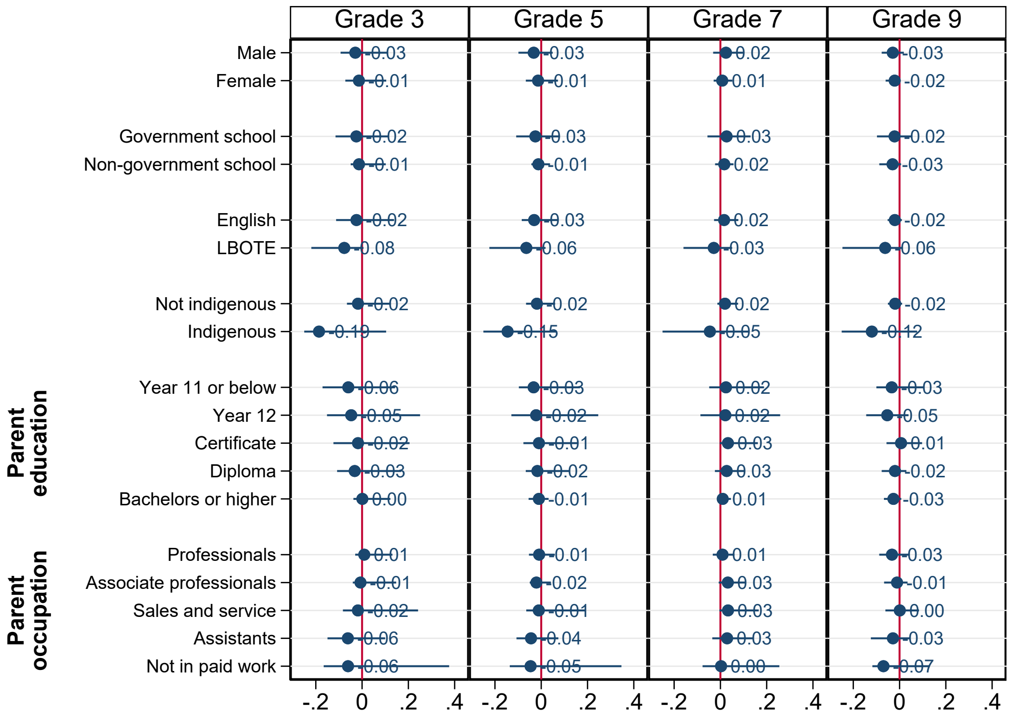
Summary: There is evidence of only modest learning losses from school closures, with no systematic pattern across grade levels of larger learning losses for any particular test component.

Figure 3: Event Study: Per 100 days School Closures



Notes: The figure shows estimated β_t coefficients for Equation (2) for the composite test score. Bar range shows 95 percent confidence interval.
Summary: The variation in test scores seen during school closures was similar to that seen in previous years.

Figure 4: Heterogeneity: Per 100 days School Closures



Notes: The figure shows estimated β coefficients for Equation (3). LBOTE is students with Language Background Other Than English. Lower confidence interval for indigenous students truncated at -0.25 per 100 days in figure.

Summary: Learning loss was small across most socio-economic characteristics. There is suggestive evidence of larger learning loss for indigenous students and those with a LBOTE.

Supplement (Not for Publication)

A Appendix A: HILDA Survey

This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey, conducted by the Melbourne Institute of Applied Economic and Social Research on behalf of the Australian Government Department of Social Services (DSS) (Wave 20, ADA Dataverse.) The findings and views reported in this paper, however, are those of the authors and should not be attributed to the Australian Government, the DSS, or the Melbourne Institute. The data used are available free of charge to researchers through the National Centre for Longitudinal Data Dataverse at the Australian Data Archive (https://dataverse.ada.edu.au/dataverse/ncld). Access is subject to approval by the Australian Government Department of Social Services and is conditional on signing a license specifying terms of use.

B Appendix B: Additional Results

B.1 Access to technology and study space while learning from home

We measure students' access to technology and study space during school closures using data from the Longitudinal Study of Australian Children (LSAC). The LSAC data contains responses from a representative sample of 1,300 individuals aged 16 to 17 when schools were closed. We use responses to questions in wave 9C1 which asked: "Please think about the period when restrictions were first at their peak. For most people this would have been between March and May 2020. During the coronavirus restriction period, how often did you have the following? Reliable internet access for all my needs". Similar questions were asked about access to sufficient electronic devices and study space. The five available responses were: never, rarely, sometimes, often and always. We classify an individual as having sufficient access to the internet, electronic devices or space if they answered sometimes, often or always. 96, 98 and 98 per cent of respondents reported having sufficient access to reliable internet, electronic devices and space, respectively.[9]

We also look at access across different socio-demographic groups by running the following regression sep-

[9] If we classify individuals as having access to sufficient resources if they answered often or always then 88, 95 and 90 per cent of respondents reported having sufficient access to reliable internet, electronic devices and space.

arately for access to each resource:

$$r_i = \sum_k \beta_k I_{i,k} + \varepsilon_i \tag{B.1}$$

where r_i is a dummy variable that is equal to one if individual i has sufficient access to either reliable internet, electronic devices or study space and k is the characteristic heterogeneity of interest (e.g. highest level of parental education) and I_k is a dummy variable taking the value for characteristic k, where characteristics are defined the same way as in Section 4. For each socio-demographic characteristic we test whether all the β_k coefficients are equal. Table S5 reports the associated F-statistics and p-values. We find no difference in access to reliable internet, electronic devices or study space across any of the socio-demographic characteristics or by geographic location.

The LSAC surveyed children aged 16 to 17. One concern in using the responses from this survey could be that these children might have access to better resources because they were closer to finishing school compared to younger children. However, we believe that the results from this study can generalize to younger children. Firstly, around half of all survey respondents had younger siblings living in the same house. Secondly, internet penetration rates in Australia are high—close to 90 per cent of the population has access to the internet and internet speeds are fast. Thirdly, data from Programme for International Student Assessment (PISA) is consistent with estimates from the LSAC. The PISA study which was conducted in 2018 found that 98, 98 and 88 per cent of 15 year old students in Australia had access to internet, electronic devices and a quiet place to study at home respectively. Two-thirds of students reported having access to three or more electronic devices at home.

B.2 Parental supervision of learning activities

We measure parental supervision of learning activities during school closures using data from Household, Income and Labour Dynamics in Australia (HILDA) Survey, a nationally representative survey of Australian households. We use responses from wave 20 of the survey and make use of two questions, the first, "Did children staying home from school have any impact on your ability to undertake paid work?" and the second "And what about other members of this household? Did children staying home from school have any impact on their ability to undertake paid work?" Impact on parent's ability to undertake paid work is measured by the parent having to take either paid or unpaid leave, reducing their work hours or quitting their job. Our sample consists of responses from 4,127 parents surveyed from 4 August 2020 to 7 February 2021. Table

S6 shows the breakdown of responses by household members for duration of school closures of less than 30 days (the median duration of school closures) and more than 70 days (the 90th percentile) for households where both parents worked. We find that the probability of an effect on at least one parent's ability to undertake paid work increases by 10 percentage points from 34 to 44 per cent as the length of school closure increases from less than 30 days to more than 70 days.

Table S1: Test Score Summary Statistics: Grades 3 and 5

	2013-2019				2021				2022			
	mean	p10	p50	p90	mean	p10	p50	p90	mean	p10	p50	p90
Grade 3												
NSW metro	5.96	4.69	5.98	7.19	6.09	4.79	6.12	7.32	6.07	4.72	6.09	7.37
NSW regional	5.55	4.28	5.58	6.75	5.63	4.31	5.69	6.83	5.61	4.23	5.66	6.86
VIC metro	6.01	4.84	6.01	7.18	6.06	4.88	6.08	7.23	6.05	4.78	6.06	7.30
VIC regional	5.72	4.58	5.73	6.86	5.76	4.57	5.79	6.87	5.72	4.49	5.74	6.93
QLD metro	5.76	4.50	5.79	6.98	5.85	4.57	5.90	7.04	5.79	4.43	5.82	7.04
QLD regional	5.48	4.19	5.51	6.71	5.47	4.12	5.55	6.69	5.40	4.04	5.45	6.64
WA metro	5.79	4.50	5.82	7.01	5.87	4.58	5.92	7.07	5.87	4.54	5.90	7.12
WA regional	5.27	3.85	5.36	6.57	5.39	3.89	5.51	6.63	5.38	3.91	5.45	6.68
SA metro	5.66	4.43	5.69	6.83	5.79	4.55	5.83	6.93	5.79	4.52	5.81	6.99
SA regional	5.35	4.08	5.40	6.54	5.46	4.17	5.54	6.62	5.44	4.15	5.47	6.64
TAS metro	5.65	4.32	5.69	6.92	5.76	4.43	5.80	6.97	5.71	4.32	5.76	6.99
TAS regional	5.47	4.17	5.52	6.68	5.46	4.08	5.53	6.70	5.46	4.07	5.46	6.80
ACT	5.89	4.65	5.92	7.08	5.91	4.68	5.96	7.08	5.93	4.63	5.95	7.14
NT	4.72	2.67	4.96	6.45	4.93	2.76	5.23	6.61	5.13	3.24	5.33	6.64
Total	5.78	4.50	5.81	7.02	5.87	4.56	5.92	7.10	5.85	4.47	5.87	7.14
Grade 5												
NSW metro	7.68	6.41	7.67	8.96	7.79	6.57	7.81	9.02	7.78	6.55	7.79	9.00
NSW regional	7.24	6.00	7.27	8.43	7.33	6.08	7.39	8.48	7.28	6.05	7.33	8.43
VIC metro	7.70	6.57	7.69	8.87	7.77	6.68	7.77	8.89	7.75	6.63	7.76	8.89
VIC regional	7.41	6.34	7.41	8.51	7.46	6.41	7.48	8.51	7.42	6.35	7.44	8.48
QLD metro	7.51	6.29	7.52	8.72	7.58	6.40	7.61	8.74	7.54	6.34	7.56	8.73
QLD regional	7.21	5.96	7.25	8.43	7.21	5.92	7.29	8.39	7.14	5.85	7.20	8.30
WA metro	7.53	6.31	7.55	8.75	7.64	6.44	7.67	8.79	7.61	6.43	7.65	8.77
WA regional	7.03	5.62	7.13	8.31	7.13	5.66	7.25	8.39	7.15	5.76	7.24	8.36
SA metro	7.38	6.20	7.40	8.56	7.50	6.36	7.53	8.63	7.50	6.34	7.52	8.66
SA regional	7.07	5.87	7.12	8.23	7.12	5.91	7.18	8.25	7.15	5.98	7.20	8.25
TAS metro	7.35	6.05	7.37	8.61	7.43	6.14	7.50	8.63	7.44	6.21	7.49	8.59
TAS regional	7.13	5.90	7.17	8.32	7.16	5.93	7.20	8.30	7.17	5.95	7.25	8.25
ACT	7.61	6.44	7.63	8.77	7.62	6.47	7.66	8.72	7.66	6.54	7.69	8.76
NT	6.44	4.29	6.73	8.21	6.60	4.53	6.90	8.28	6.80	4.76	7.09	8.33
Total	7.50	6.26	7.51	8.74	7.58	6.36	7.62	8.77	7.56	6.34	7.59	8.76

Notes: The table shows standardized NAPLAN scores by region for the periods 2013-2019, 2021 and 2022 for Grades 3 and 5 students. Scores are standardized by the grade-level national standard deviation over the period 2013-2019.

Table S2: Test Score Summary Statistics: Grades 7 and 9

	2013-2019				2021				2022			
	mean	p10	p50	p90	mean	p10	p50	p90	mean	p10	p50	p90
Grade 7												
NSW metro	8.24	6.98	8.22	9.55	8.31	7.02	8.32	9.61	8.33	7.05	8.34	9.63
NSW regional	7.78	6.57	7.81	8.99	7.77	6.45	7.83	9.00	7.82	6.52	7.87	9.04
VIC metro	8.24	7.09	8.22	9.43	8.28	7.13	8.28	9.47	8.31	7.12	8.32	9.51
VIC regional	7.91	6.81	7.91	9.02	7.90	6.77	7.92	9.03	7.90	6.71	7.93	9.04
QLD metro	8.12	6.93	8.12	9.32	8.14	6.90	8.16	9.38	8.14	6.90	8.15	9.38
QLD regional	7.78	6.58	7.81	8.98	7.75	6.43	7.80	9.02	7.72	6.41	7.77	8.98
WA metro	8.16	6.94	8.18	9.39	8.22	6.94	8.26	9.44	8.23	7.00	8.26	9.46
WA regional	7.69	6.31	7.77	8.96	7.71	6.28	7.81	8.95	7.69	6.28	7.79	8.97
SA metro	8.09	6.92	8.09	9.27	8.13	6.90	8.16	9.33	8.09	6.86	8.11	9.30
SA regional	7.78	6.65	7.81	8.93	7.80	6.59	7.85	8.96	7.73	6.52	7.77	8.90
TAS metro	7.95	6.70	7.97	9.19	7.97	6.61	8.03	9.18	8.02	6.73	8.06	9.26
TAS regional	7.70	6.50	7.73	8.88	7.63	6.25	7.69	8.90	7.74	6.43	7.81	8.94
ACT	8.23	7.06	8.25	9.41	8.21	6.99	8.27	9.37	8.22	7.04	8.27	9.38
NT	7.05	4.89	7.34	8.85	7.24	5.05	7.51	8.95	7.34	5.24	7.61	8.97
Total	8.08	6.86	8.09	9.33	8.12	6.84	8.15	9.38	8.13	6.85	8.16	9.40
Grade 9												
NSW metro	8.60	7.36	8.60	9.88	8.58	7.34	8.61	9.82	8.62	7.37	8.65	9.83
NSW regional	8.16	6.96	8.19	9.36	8.10	6.83	8.19	9.28	8.12	6.89	8.19	9.30
VIC metro	8.58	7.44	8.58	9.76	8.55	7.42	8.57	9.69	8.59	7.41	8.63	9.73
VIC regional	8.28	7.18	8.29	9.41	8.22	7.11	8.27	9.34	8.26	7.12	8.31	9.38
QLD metro	8.44	7.28	8.45	9.62	8.44	7.21	8.49	9.64	8.42	7.16	8.48	9.64
QLD regional	8.11	6.93	8.14	9.31	8.04	6.72	8.13	9.27	8.04	6.73	8.11	9.28
WA metro	8.60	7.42	8.63	9.78	8.65	7.55	8.68	9.74	8.64	7.51	8.67	9.77
WA regional	8.16	6.89	8.24	9.38	8.22	6.93	8.33	9.38	8.17	6.88	8.27	9.36
SA metro	8.41	7.25	8.44	9.59	8.43	7.25	8.48	9.58	8.42	7.21	8.48	9.60
SA regional	8.09	6.96	8.13	9.22	8.14	6.99	8.23	9.22	8.10	6.91	8.18	9.19
TAS metro	8.29	7.05	8.32	9.52	8.29	6.97	8.40	9.46	8.27	7.00	8.35	9.45
TAS regional	8.05	6.86	8.08	9.18	8.06	6.79	8.12	9.21	7.94	6.66	8.04	9.13
ACT	8.61	7.42	8.65	9.79	8.54	7.31	8.63	9.66	8.55	7.34	8.63	9.67
NT	7.58	5.48	7.85	9.28	7.76	5.86	7.99	9.30	7.90	6.14	8.07	9.38
Total	8.44	7.24	8.46	9.68	8.43	7.20	8.48	9.63	8.45	7.20	8.50	9.66

Notes: The table shows standardized NAPLAN scores by region for the periods 2013-2019, 2021 and 2022 for Grades 7 and 9 students. Scores are standardized by the grade-level national standard deviation over the period 2013-2019.

Table S3: Means by State

	Female	Non-Govt.	LBOTE	Bachelors	Professional	Indigenous
NSW metro	0.49	0.37	0.38	0.44	0.54	0.04
NSW regional	0.49	0.32	0.05	0.26	0.40	0.13
VIC metro	0.49	0.38	0.36	0.46	0.52	0.01
VIC regional	0.49	0.36	0.06	0.29	0.44	0.04
QLD metro	0.49	0.35	0.16	0.38	0.50	0.05
QLD regional	0.49	0.30	0.09	0.23	0.36	0.14
WA metro	0.49	0.37	0.24	0.40	0.52	0.04
WA regional	0.49	0.26	0.12	0.21	0.35	0.17
SA metro	0.49	0.40	0.20	0.37	0.49	0.03
SA regional	0.48	0.25	0.05	0.20	0.37	0.08
TAS metro	0.49	0.36	0.07	0.29	0.45	0.08
TAS regional	0.49	0.18	0.03	0.17	0.34	0.14
ACT	0.49	0.43	0.24	0.55	0.64	0.03
NT	0.49	0.29	0.42	0.24	0.35	0.42
Total	0.49	0.36	0.24	0.38	0.49	0.06

Notes: The table shows means over the period 2013-2019 for students in Grades 3, 5, 7 and 9. *Non-Govt.* is non-government school, *LBOTE* is language background other than English, *Bachelors* is an indicator for either parent having a Bachelors degree or higher, *Professional* is an indicator for either parent's occupation being professional or associate professional, and *Indigenous* is Aboriginal or Torres Strait Islander.

Table S4: School Attendance Rates by State

	Pre-pandemic period				Pandemic period				
	2019-T1	2019-T2	2019-T3	2019-T4	2021-T1	2021-T2	2021-T3	2021-T4	2022-T1
NSW	89.7	91.3	92.3	93.5	88.3	90.4	91.8	93.3	83.4
VIC	91.2	91.9	92.5	93.3	91.0	91.9	92.8	93.9	86.3
QLD	89.3	90.9	91.8	93.0	87.6	89.7	91.0	92.5	83.9
WA	88.1	90.7	91.8	93.0	86.6	89.7	91.0	92.4	82.3
SA	89.1	91.1	92.1	93.1	87.8	90.3	91.6	92.9	83.2
TAS	89.4	90.9	91.7	92.5	87.5	89.5	90.6	91.7	83.1
ACT	89.4	90.5	91.1	91.7	90.1	91.1	91.7	92.2	84.0
NT	69.7	86.7	88.7	90.3	68.6	87.0	89.0	90.6	65.1
Total	87.0	90.5	91.5	92.6	85.9	90.0	91.2	92.4	81.4

Notes: The attendance rate is the number of actual full-time equivalent student-days attended by full-time students in Years 1-10 as a percentage of the total number of possible student-days attended (ACARA, 2022*b*). Data are averages by school term for students in Grades 1-10. Gray shading denotes that schools were closed for at least half term days. Data are unavailable for 2020 and by region for each school term.

Summary: School attendance was high when schools were closed and lessons were conducted online.

Table S5: Heterogeneity: Sufficient Access to Resources to Learn From Home
F-statistics for test that coefficients are equal, p-values in brackets

	Reliable Internet	Electronic Devices	Study Space
Gender	0.534	0.000	0.024
	(0.465)	(0.996)	(0.876)
Govt/Non-govt school	0.040	1.470	0.225
	(0.842)	(0.226)	(0.635)
Language background	0.134	0.003	0.001
	(0.714)	(0.957)	(0.978)
Parent education	0.223	1.318	0.966
	(0.926)	(0.261)	(0.425)
Parent occupation	0.593	1.381	0.453
	(0.620)	(0.247)	(0.715)
Indigenous status	0.613	0.313	0.440
	(0.434)	(0.576)	(0.507)
Metro/Non-metro location	1.698	2.276	2.046
	(0.193)	(0.132)	(0.153)
State	0.852	0.525	0.880
	(0.530)	(0.790)	(0.509)

Notes: We test whether there are difference in access to reliable internet, electronic devices and space to learn from home by socio-demographic characteristics. The table reports F-statistics and p-values in parentheses associated with test that the β_k coefficients in Equation (B.1) are jointly equal.

Summary: There was little difference in access to resources across states and socio-demographic groups.

Table S6: Did School Closures Affect Parent's Ability to Undertake Paid Work
Percentage of Response by Category

School closure length: <30 days
(Median length of school closures)

		Parent 1 Yes	Parent 1 No
Parent 2	Yes	10.0	19.6
	No	3.9	66.5

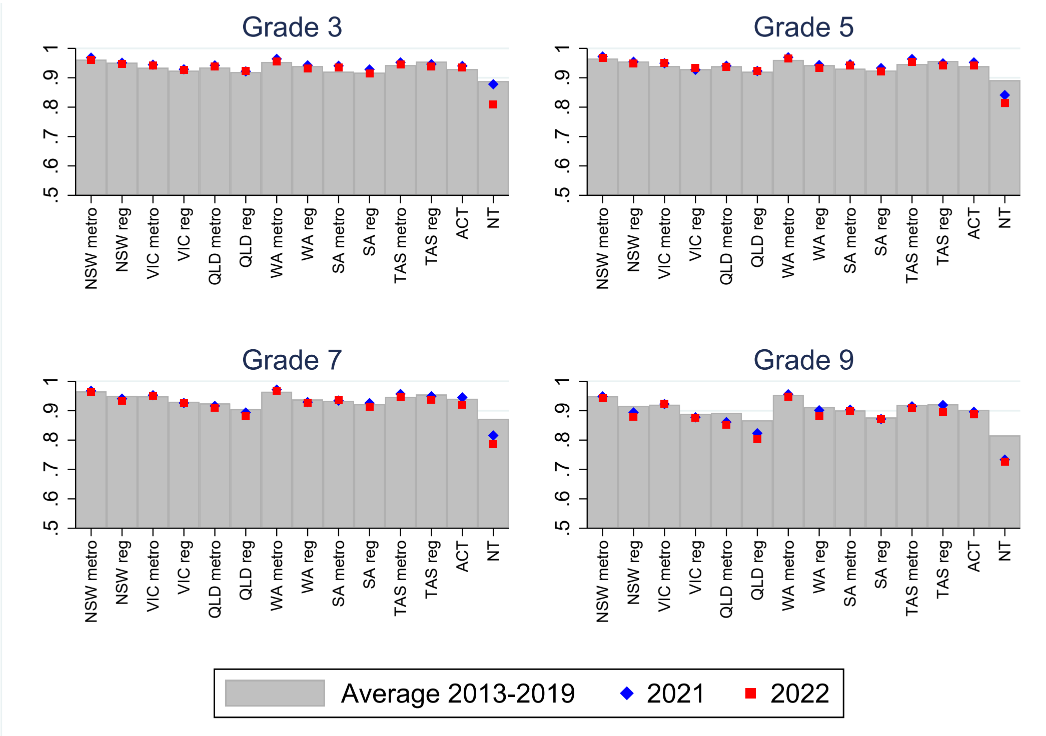
School closure length: >70 days
(90th percentile of school closure length)

		Parent 1 Yes	Parent 1 No
Parent 2	Yes	14.9	23.6
	No	4.1	57.3

Notes: The percentage of responses from parents by category to the question: "Did children staying home from school have any impact on your ability to undertake paid work?". Conditional on both parents initially being in paid employment. Impact on the ability of parents to undertake paid work is measured by parents having to take paid or unpaid leave, reducing their work hours or quitting their job. The data is sourced from HILDA Wave 20 Household Questions 33, 35-38.

Summary: The probability that there was a reduction in the ability of at least one parent to undertake paid work increased by close to 10 percentage points when the duration of learning from home increased from 30 days (the median amount of time schools were closed for) to 70 days (the 90th percentile).

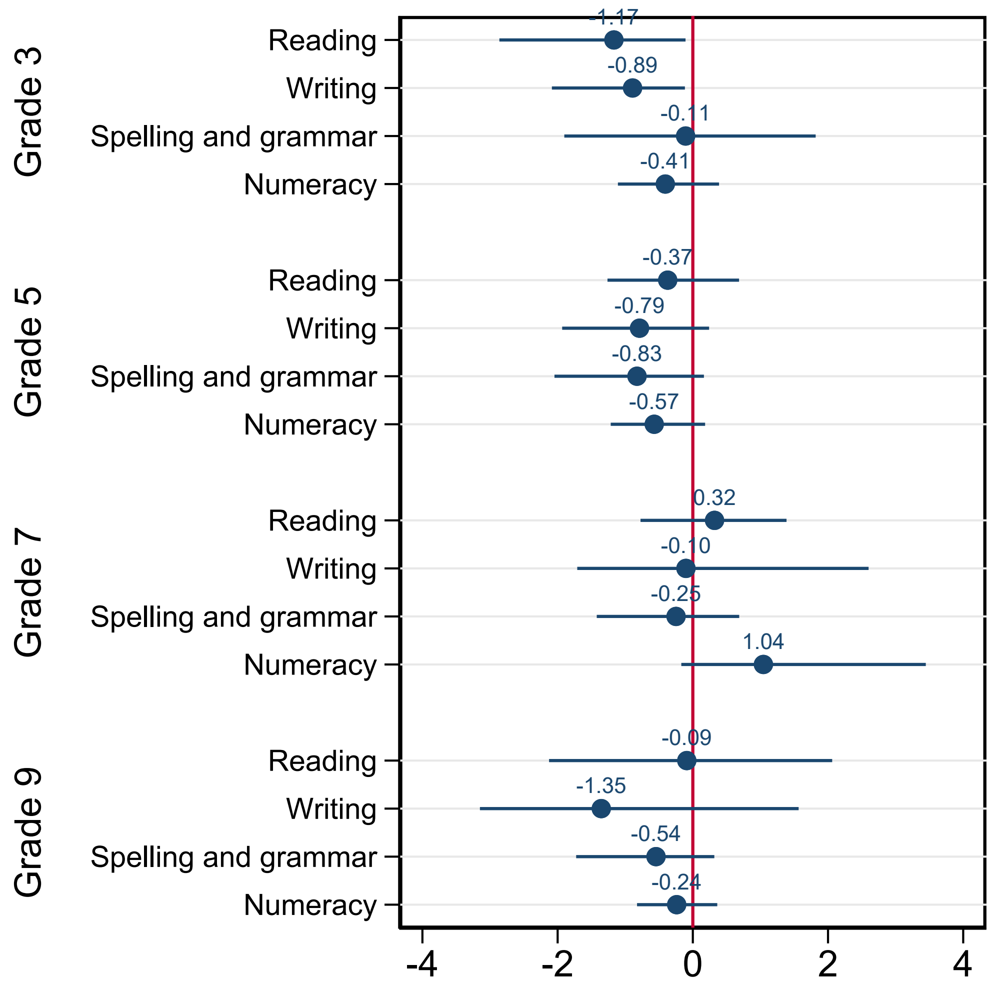
Figure S1: Participation Rates by Region



Notes: Participation is the fraction of students taking a NAPLAN test, by region and grade level. Exempt students (those with a significant disability and migrants within the past year from a non-English speaking country) are excluded from the calculation of participation rates.

Summary: Participation rates remained high in 2021 and 2022.

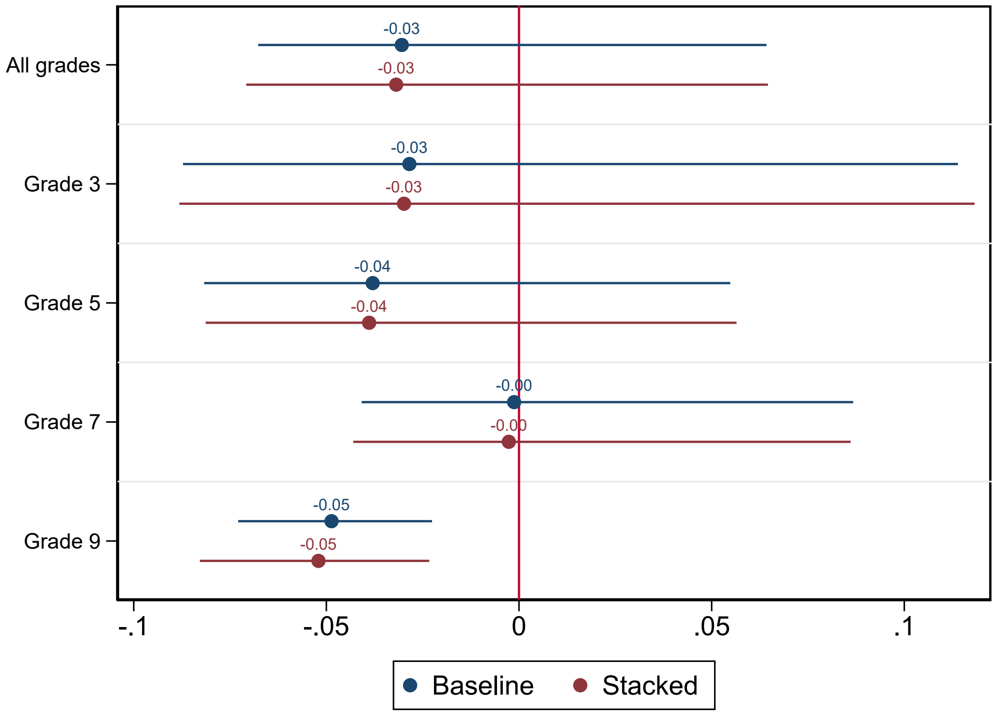
Figure S2: Meeting Minimum Standards: Per 100 days School Closures



Notes: The figure shows estimated β coefficients for Equation (1) where the dependent variable is replaced by an indicator equal to 1 if the student is meeting minimum national standards at their grade level. Units are percentage points.

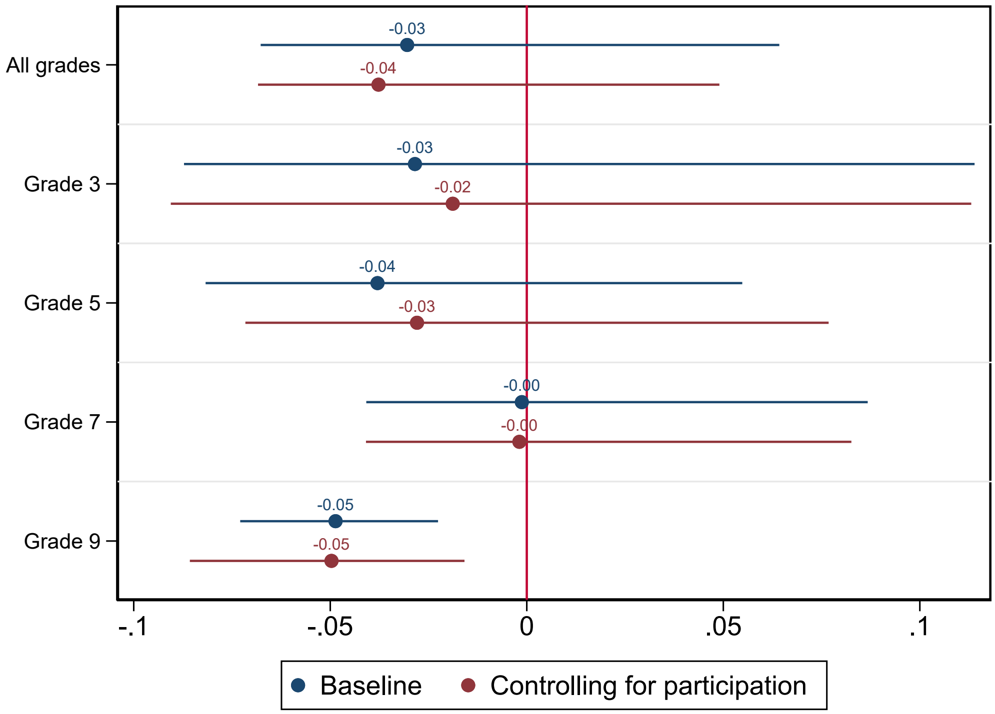
Summary: School closures did not have a large effect on achievement of minimum standards.

Figure S3: Stacked Difference-in-Difference: Per 100 days School Closures



Notes: The figure shows estimated β coefficients for Equation 4, the stacked difference-in-difference model.
Summary: The baseline and stacked difference-in-difference results are very similar.

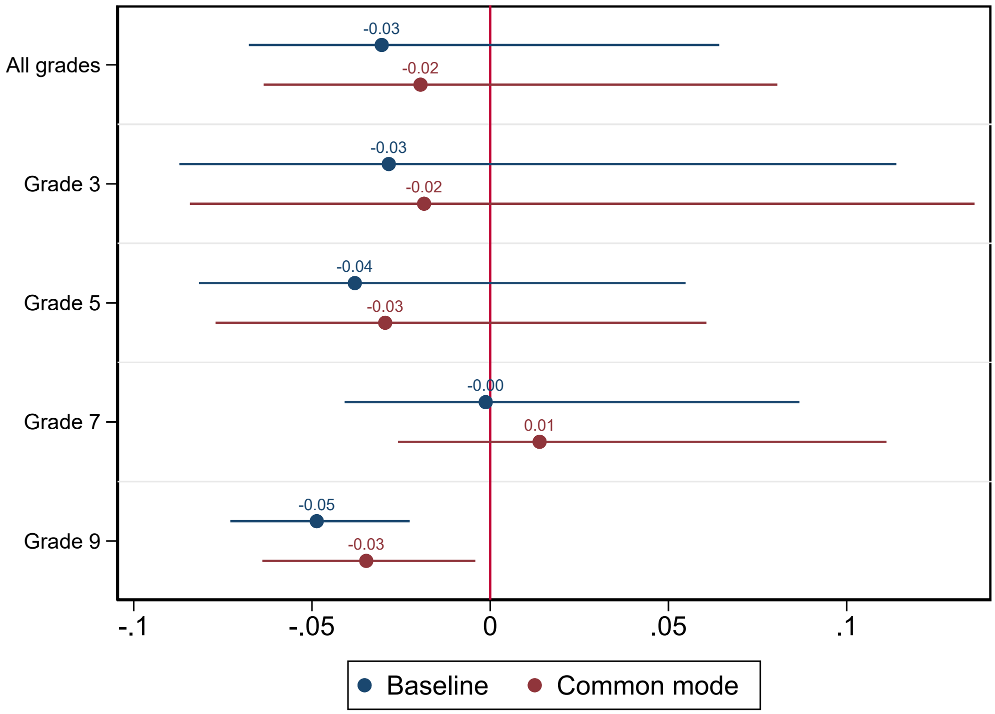
Figure S4: Baseline Results Including Participation as a Control: Per 100 days School Closures



Notes: The figure shows estimated β coefficients for Equation (1) for the main specification and for a specification that includes mean participation rate for each region and year as an additional control variable.

Summary: Including the region-level participation rate as an additional control variable has a negligible effect on the results.

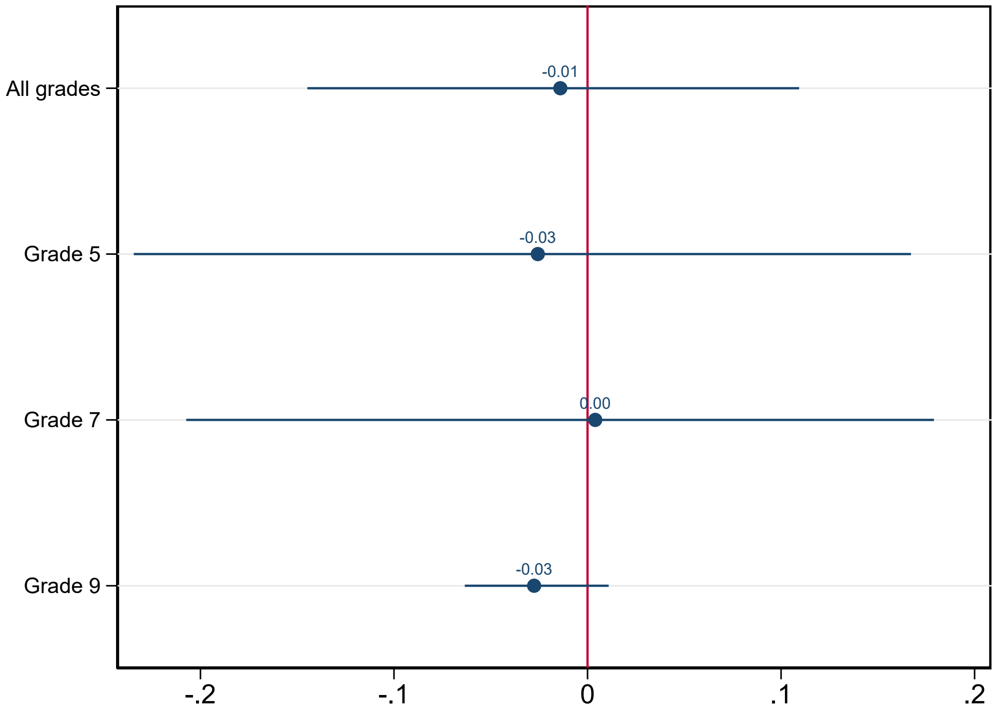
Figure S5: Baseline Results by Test Mode: Per 100 days School Closures



Notes: The figure shows estimated β coefficients for Equation (1) for the main specification and for the subset of students where test mode was the same in each year across states. There was a transition from paper-based to online testing in the period 2018-2021. The *Common mode* specification restricts the sample to students with the most common test mode in each year: paper-based 2013-2018 and online for 2019, 2021 and 2022.

Summary: Restricting the sample to online testing from 2019 has little effect on the results.

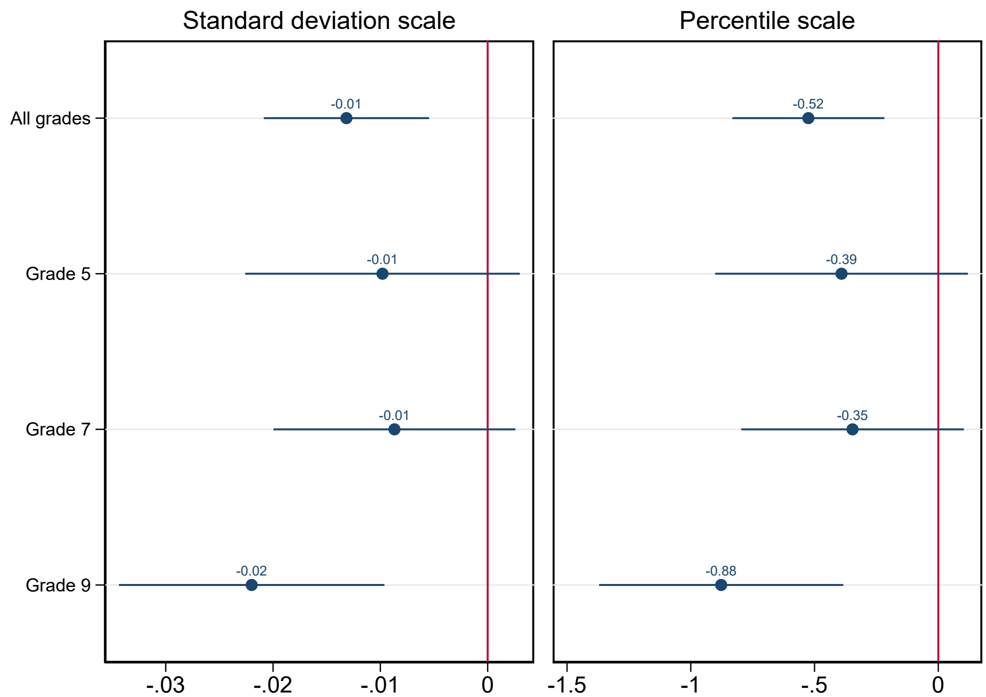
Figure S6: Change in NAPLAN Score: Per 100 days School Closures



Notes: The figure shows estimated β coefficients for Equation (1) where the dependent variable is replaced by the student-level change in score from NAPLAN tests taken 2 years prior. The change in score is unavailable for 2022 because the 2020 NAPLAN was canceled. Students take the NAPLAN for the first time in Grade 3 so no previous score is available for Grade 3 students.

Summary: Point estimates are similar when the dependent variable is the change in NAPLAN score. Standard errors are wide because 2022 data are excluded and there is less region-level variation in school closures in 2021.

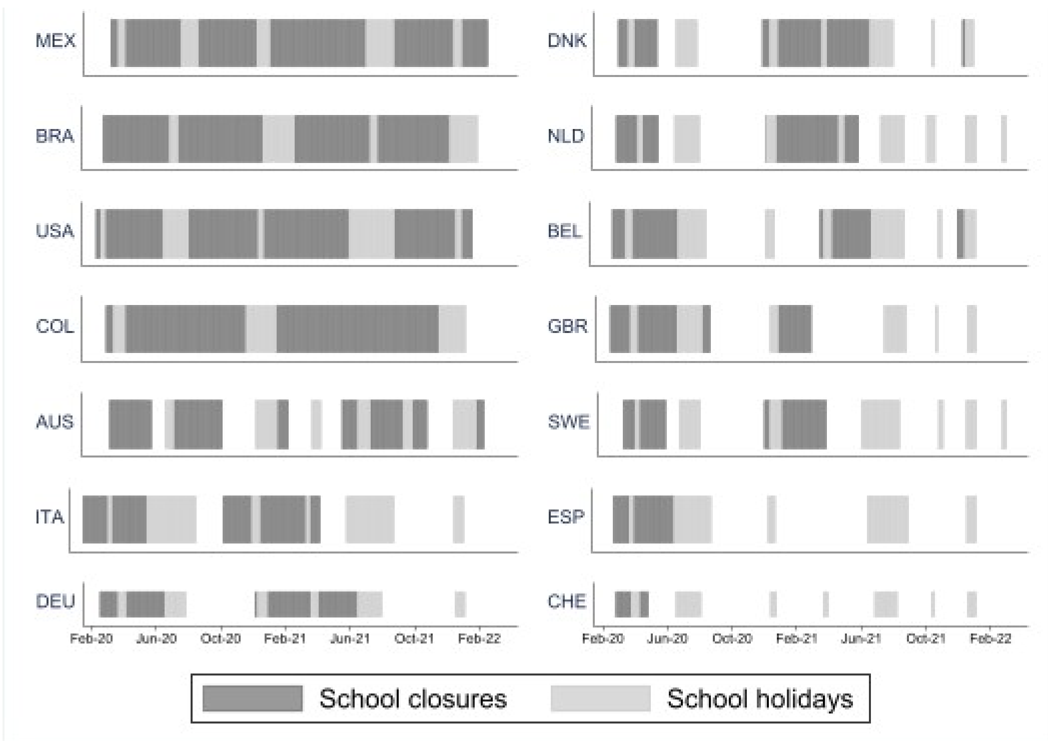
Figure S7: Baseline results: Before-After Learning Trajectories Methodology



Notes: The figure reports results using the difference-in-difference model specification from Equation (5). Estimates are for students in Victoria 2013-2019 and 2021. Standard errors are clustered at the school level following Engzell, Frey and Verhagen (2021). Percentile estimates in the right panel are computed by transforming the estimates using Equation (3) in Engzell, Frey and Verhagen (2021).

Summary: Results are similar using the learning trajectories methodology.

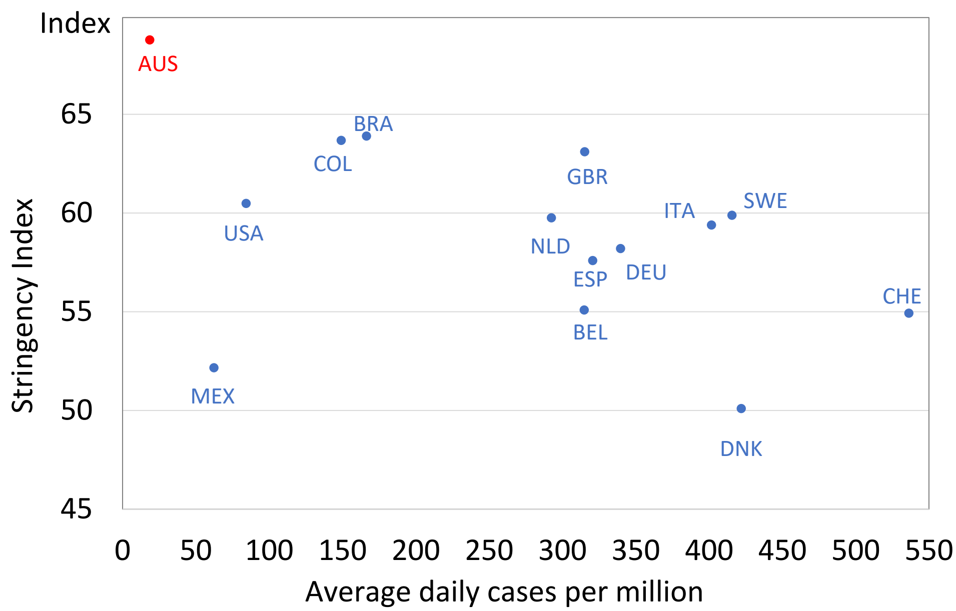
Figure S8: Duration of School Closures by Country



Notes: The figure shows the length of school closures by countries in the meta-analysis of Betthäuser, Bach-Mortensen and Engzell (2023). Countries are ranked from those with the longest closures to those with the shortest. The data is sourced from UNESCO (2022). The dark shaded areas represent school closures in at least one part of the country.

Summary: The length of school closures in Australia was close to the median and longer than that in European countries.

Figure S9: Restrictions on Human Movement and COVID-19 Cases During School Closures



Notes: The figure shows the strictness of policies to restrict human movement, as measured by the Oxford Coronavirus Government Response Tracker (Mathieu et al., 2021), and average daily COVID-19 case numbers during periods of school closures by country. Higher values of the Stringency Index reflect stricter policies restricting human movement. The countries are those referred to in the meta-analysis of Betthäuser, Bach-Mortensen and Engzell (2023).

Summary: Australia had the strictest policies restricting human movement and the lowest average of daily COVID-19 cases during school closures.