PIER

PUEY UNGPHAKORN INSTITUTE
FOR ECONOMIC RESEARCH

Learning Losses from School Closure Due to the COVID-19 Pandemic for Thai Kindergartners

by

Weerachart T. Kilenthong, Khanista Boonsanong,
Sartja Duangchaiyoosook, Wasinee Jantorn and Varunee Khruapradit

March 2022
Discussion Paper
No. 174

Learning Losses from School Closure due to the COVID-19 Pandemic for Thai Kindergartners*

Weerachart T. Kilenthong, Khanista Boonsanong,
Sartja Duangchaiyoosook, Wasinee Jantorn and Varunee Khruapradit
Research Instititute for Policy Evaluation and Design (RIPED)
University of the Thai Chamber of Commerce

March 10, 2022

Abstract

This paper presents empirical evidence of learning losses from school closure due to the COVID-19 pandemic for kindergartners using a large-scale school readiness survey in Thailand. Its findings indicate that school closure during the outbreak of COVID-19 causes enormous learning losses in cognitive skills, especially in mathematics and working memory. The negative impact is heterogeneous across several dimensions, including child gender, special needs, wealth, having private tutoring, caregiver's education and parental absence. This paper also estimates daily learning gains, of which significant results confirm that going to school has significantly benefited young children, especially in receptive language, mathematics and working memory.

*We would like to thank Warn N. Lekfuangfu, Suphanit Piyapromdee and Nada Wasi for helpful comments. The authors are grateful to financial support from the Equitable Education Fund (EEF). The findings and conclusions contained in the report are those of the author(s) and do not necessarily represent the views of the funders. Corresponding author: Weerachart T. Kilenthong. Email: tee@riped.utcc.ac.th (Weerachart T. Kilenthong); serene@riped.utcc.ac.th (Khanista Boonsanong); kei@riped.utcc.ac.th (Sartja Duangchaiyoosook); wasinee_jun@riped.utcc.ac.th (Wasinee Jantorn); varunee@riped.utcc.ac.th (Varunee Khruapradit)

Keyword: learning losses; school closure; school readiness; daily learning gains; COVID-19; COVID slide; disadvantaged children; working memory; math; literacy; receptive language; non-cognitive

JEL Code: I21; I25; J24.

1 Introduction

School closure has been a standard policy measure against the COVID-19 pandemic around the world. Its immediate benefit in reducing the risk of infection is readily realized but the cost incurred on human capital accumulation is not obvious. There is limited empirical evidence on learning losses due to school closure during the pandemic with the exception of Tomasik et al. (2020); Ardington et al. (2021); Contini et al. (2021); Engzell et al. (2021); Halloran et al. (2021); Lewis et al. (2021); Maldonado and De Witte (2021); Schult and Lindner (2021), all of which studied the effects on either primary or secondary students. So far, there is no empirical evidence of the impact on younger children even though this is a critical period for human capital development (e.g., Knudsen et al., 2006; Cunha et al., 2010; Currie and Almond, 2011).

To the best of our knowledge, this paper is the first paper to provide empirical evidence of learning losses from school closure due to the COVID-19 pandemic for kindergartners. We take advantage of a localized lockdown policy of the Thai government, under which only 28 out of 77 provinces were lockdowned between January and February 2021 to counter the second wave of the outbreak. This localization policy creates provincial variation in school closure. Importantly, we consider the provincial lockdown as a natural experiment, a key instrument to identify learning losses.

On the other hand, the Thailand school readiness survey was implemented in 25 provinces at the beginning of 2021 after the lockdown was lifted. Our identification relies on the fact that 4 out of 25 surveyed provinces, namely Kanchanaburi, Nonthaburi, Phra Nakhon Si Ayutthaya and Samut Prakan, were lockdowned. Timing of the survey (after the lockdown was lifted) is also key to our ability to capture the impact of school closure since it creates a significant variation of in-person schooling days across provinces. In addition, this so-called 2021 Thailand school readiness survey collected information about school closure due to the outbreak and number of days closed. The latter piece of information allows us to estimate the daily learning gains. The data also reveal non-compliance regarding school closure; that is, many schools in no-lockdowned provinces were closed while some in lockdowned provinces were not.

Technically, this paper deals with the non-compliance problem using the provincial lockdown and the accumulated number of COVID confirmed cases in each province as instruments. These instruments should also be useful for dealing with potential endogeneity problems, e.g., teachers in lower-performing schools may be more likely to put in less effort and, therefore, more likely to switch to remote learning mode. Our key identification assumption is that the outbreak of COVID-19 is not correlated with any unobserved factors affecting child development. Weak instrument, under-identification and over-identification tests were performed to check the validity of the instruments.

To rule out spurious relationships or other unobserved variations that could be driving the results over and above school closure, we exploit the 2020 Thailand school readiness survey, which was conducted right before the first wave of COVID-19 pandemic in Thailand (started in January 2020 and ended in March 2020). This survey allows us to perform placebo or falsification tests. These tests should help ensure that our benchmark models are not prone to false positives. Indeed, the test results confirm that we should not be worried that unobserved heterogeneity may contaminate our results.

This paper contributes to the current and future policy debates regarding school interruption due to natural disasters or pandemics by providing an empirical evidence of the cost of school closure for kindergartners. Our results confirm that school closure during the outbreak of COVID-19 causes enormous learning losses in cognitive skills, especially in mathematics and working memory. This is, of course, only one side of the story, the cost of school closure. We still need more evidence for the benefit of school closure or risks associated with school opening during the pandemic (as shown in Isphording et al., 2021) to produce an optimal response to the outbreak of COVID-19 or a future pandemic. It is also worthy of emphasis that going to school has significantly benefited young children, especially in receptive language, mathematics and working memory.

In addition, the negative impact, called COVID slide, is heterogeneous across groups. We found that the negative impact is worse of more advantaged children. The outbreak of COVID-19 has put a physical barrier to not just formal schooling but also other means, e.g., private tutoring. This is distinct from the summer slide literature (Alexander et al., 2001, 2007), which found more negative effects on more disadvantaged children because, their limited resources constrain them from getting access to extra activities like private tutoring or summer camps. However, this does not mean that school closure due to COVID-19 did not affect disadvantaged children. An optimal public policy should focus on how to help disadvantaged children recovering their losses once the outbreak is over because the speed of recovery may depend on resources the family has.

This paper belongs to the literature on learning losses from school closure due to COVID-19 (Tomasik et al., 2020; Ardington et al., 2021; Contini et al., 2021; Engzell et al., 2021; Halloran et al., 2021; Lewis et al., 2021; Maldonado and De Witte, 2021; Schult and Lindner, 2021). All of the papers dealt with learning losses of either primary or secondary students. Our paper is the first to estimate learning losses due to the COVID-19 pandemic for kindergartners. Another advantage of our paper is the richness of children and household information, which allows us to control for more individual characteristics. In addition, we can estimate and find interesting heterogeneous effects based on the children and household information.

This paper is also related to the literature on the negative impact of school interruption due to natural disasters, wars, or teacher strikes (e.g., Ichino and Winter-Ebmer, 2004; Hansen, 2011; Sacerdote, 2012; Goodman, 2014; Cattan et al., 2017; Jaume and Willén, 2019). Our paper is clearly related to this group of research in that school interruption is caused by exogenous shocks or natural experiments. In addition, this set of research, including ours, not only provides empirical evidence for the negative impact of school interruption but also the daily learning gains from going to school. In other words, our paper provides an empirical evidence on the effect of schooling on human capital formation (e.g., Todd and Wolpin, 2003, 2007; Cunha and Heckman, 2007, 2008; Cunha et al., 2010; Del Boca et al., 2014). That is, this paper is also contributing to literature on human capital production function .

The remainder of this paper is organized as follows. Section 2 describes the situation of the outbreak of COVID-19 and school closure in Thailand. Data and estimation methods are explained in section 3. Section 4 presents empirical results of the benchmark models while section 5 discusses the heterogeneous effects. Robustness checks with respect to child, parent, household and school characteristics are presented in section 6 while section 7 covers the placebo tests. Section 8 summarizes key findings, proposes policies to facilitate learning recovery, and discusses its weaknesses. Appendix A and B contain additional tables and figures.

2 Background on the Outbreak of COVID-19 and School Closure in Thailand

The first case of COVID-19 in Thailand was confirmed on January 13, 2020. The country was then under national-wide lockdown from March 26, 2020 to April 30, 2020. Fortunately, all schools were already closed for the summer break at the time.[1] There should be no schooling effect from the first wave of the outbreak. The government later decided to move the beginning of the first semester of schooling from May 18, 2020 to July 1, 2020. This postponement was for the whole country. Some schools could not reopen or have to implement an alternating schedule, under which each half of the class alternately comes to school. This reopening delay happened mostly in Bangkok, which is not part of our sample. However, all schools were reopened in August 3, 2020. In addition, schools still had time to extend the semester then. Therefore, the variation of school days in the first semester should be minimal. Nevertheless, school days in our analysis include school days from both semesters.

The second wave of the outbreak between December 2020 and February 2021, which was part of the second semester, is key to this paper. The outbreak began on December 17, 2020 in Samut Sakhon province (not part of the 2021 survey) and later spread to many provinces and the government had to (locally) lockdown 28 out of 77 provinces on January 3, 2021. The lockdown lasted until February 7, 2021 except for three provinces including Bangkok, Samut Sakhon and Tak, all of which are not part of the 2021 survey. The local-lockdown policy, which is unique to Thailand at that time, helps generate our key variable, called provincial lockdown. Figure B.1 shows lockdowned provinces in red.

[1] In the normal time, the first semester runs from the middle of May until the beginning of October while the second semester runs from the beginning of November until the end of March.

3 Data, Measurement and Methodology

3.1 Data

The main analysis in this paper uses the 2021 Thailand school readiness data[2], collected during February to April 2021. This round is the second large-scale survey of kindergartners from both public and private schools covering 25 provinces across the country while the first round, surveyed in 2020, covered 19 provinces all of which were not included in the second round. See figure B.3 and B.2 for sampled provinces in the first and the second round. Importantly, four out of 25 surveyed provinces (Kanchanaburi, Nonthaburi, Phra Nakhon Si Ayutthaya and Samut Prakan) were severely hit by COVID-19 pandemic and lockdowned during January to February 2021.

Survey instruments for the school readiness survey comprise three main parts: child development assessment, teacher-school questionnaire, and household questionnaire, all of which are adopted mainly from the MELQO: Measuring Early Learning Quality and Outcomes (UNICEF et al., 2017) with additional questions regarding household structure, asset holdings and COVID-19 related questions.

This paper analyzes both cognitive and non-cognitive skills for children. Cognitive skills were assessed by a direct assessment method.[3] On the other hand, non-cognitive skills were derived from teacher and parent self-filled questionnaires, which could be less accurate than the cognitive skills. Therefore, we should interpret the results for non-cognitive skills more carefully. Control variables are taken from both teacher-school and household questionnaires, which are self-filled as well.

The survey design is a stratified-random sampling. First, 25 out of 58 provinces, which were not surveyed in the first round, were chosen based mainly on operational reasons. For each province except Phuket where there are only three districts, five districts (called amphoes in Thai) were randomly chosen by dividing all of them into five groups, one is the central district (called amphoe Mueang in Thai), and the other four are ranked

[2]Thailand school readiness survey is a collaboration between the Equitable Education Foundation (EEF) and Research Institute for Policy Evaluation and Design (RIPED). This survey aims to have a representative sample of Thai kindergartners in all provinces of Thailand before the end of 2022. The third round covering the remaining 33 provinces will be implemented at the beginning of 2022.

[3]To control the assessment quality, the team has developed and implemented the assessment tools using an online platform.

and equally divided using their poverty level. We then randomly chose one district from each group. For each district, seven schools were randomly chosen by dividing all schools with kindergarten classes into three groups, based on their size: small school (with number of kindergartners less than the 64th percentile); medium school (with number of kindergartners more than the 64th but less than 86 percentile); and large school. We then randomly chose one classroom for small, one classroom for medium, and up to two classrooms for large schools[4]. For each classroom, up to 15 students were randomly chosen. If a school rejected our survey request, we would randomly choose a same-size school from the same district to replace it. But, if there was no school of the same size available, we would choose from a one-step-smaller size school. If there were still fewer than 470 students, we would randomly chose a small school from the five districts one by one until the number of sampled students exceeds 470.

In principle, there should be at least 35 sampled schools and at least 470 students in each province but, unfortunately, there were five provinces (two were lockdowned while the other three were not) with fewer than 35 schools and three provinces (one was lockdowned while the other two were not) with fewer than 470 students. Several schools, especially in COVID-severely-hit provinces, rejected our survey requests. Overall, the rejection rate was 16 percent (171 out of 1,098 schools), resulting in the sample of 927 schools in total. All provinces have at least 400 students nonetheless and the average number of students is 494 per province. In total, there are 12,345 kindergartners in the whole sample.

The original sample consists of 11,478 kindergartners[5] from 25 provinces, four of which were lockdowned during January to February 2021 (including Nonthaburi, Kanchanaburi, Samut Prakan and Phra Nakhon Si Ayutthaya provinces). However, 3,673 observations have to be excluded due to missing relevant school or household information. In sum, the final baseline sample consists of at most 7,805 kindergartners. Summary statistics of key variables is reported in table A.2.

[4]There was only one classroom for some large schools, and that classroom will be chosen automatically.

[5]There were 17 children taken from a school which is not supposed to be in the sample, 13 children without age information, 605 children were either younger than 60 months old or older than 84 months old, and 232 children tested too early (no more than 10 ten school days after the latest reopening).

3.2 Measurements

3.2.1 Outcome Variables: Cognitive and Non-cognitive Skills as Latent Factors

In order to reduce measurement errors in outcome variables, we treat outcome variables, both cognitive and non-cognitive skills, as latent variables (as in, Cunha and Heckman, 2008; Cunha et al., 2010; Attanasio et al., 2020). We implement this approach in five steps as follows.

First, we generate raw scores for all items from survey data. There are 10 items for cognitive skills, including Thai letter identification, English letter identification, word reading, receptive spatial vocabulary, listening comprehension, number comparison, producing a set, mental transformation, forward and backward digit spans. Raw scores for the first 8 items are derived using two-parameter item response theory (IRT) while the last two are simply the maximum digit that a child can answer correctly. On the other hand, there are 11 items for non-cognitive skills, five of which are from the teacher questionnaire (adopted from the strengths and difficulties questionnaire: SDQ) and six of which are from the parent questionnaire (adopted from the Behavioral Problem Index: BPI).[6] The first part includes emotional symptoms, conduct problems, hyperactivity/inattention, peer-relationship problems and prosocial behaviour while the second part includes anxiousness/depression, headstrong, antisocial, hyperactive, dependent and peer problems. The original score for each question is transformed in such a way that a higher score means less problematic behavior. For example, "Break things deliberately" from BPI (scoring: Never = 2, Sometimes = 1, Always = 0) and "Rather solitary, prefer to play alone" from SDQ (scoring: Not true = 2, Somewhat true = 1, True = 0). A raw score for an item is simply the average of all questions for that item. See online appendix A for the assessment tools and questions.

Second, we derive age-standardized scores for all items using Kernel-weighted Local Polynomial Smoothing up to the third degree polynomial[7] (as in, Attanasio et al., 2020). This approach gives a standardized score with mean zero and standard deviation one

[6]The Thailand school readiness survey selectively chose to include only 16 out of 28 BPI questions.

[7]There are few observations whose predicted variance of an item is negative, and, therefore, their age-standardized scores for that item will be missing values. The maximum number of observations encountering this problem is 15 observations, which is infinitesimal relative to the sample size.

for each age group. Let Y_j^s be the standardized score of item j for skill s. See online appendix B for detailed derivation.

The third step is to apply an exploratory factor analysis (EFA) using age-standardized scores to group items together (following Heckman et al., 2013; Attanasio et al., 2020). See online appendix C for the details. This process leads to six groups of items, namely literacy (Thai letter identification, English letter identification, word reading,), receptive language (listening comprehension, receptive spatial vocabulary), math (mental transformation, number comparison, producing a set), working memory (backward and forward digit spans), non-cognitive-SDQ (conduct problems, emotional symptoms, hyperactivity/inattention, prosocial behaviour) and non-cognitive-BPI (headstrong, anxiousness/depression, antisocial, hyperactive, dependent, peer problems).[8]

The fourth step is to estimate the following dedicated measurement system or factor model in which each item only proxies one latent factor.

$$Y_j^s = \alpha_j^s + \lambda_j^s \theta^s + \varepsilon_j^s, \text{ for } j = 1 \ldots, J_s \tag{1}$$

where Y_j^s is an age-standardized score of item j for latent factor s, θ^s is a latent factor s, λ_j^s is a factor loading of item j for factor s, and ε_j^s is a mean zero measurement error term which is assumed to be independent of the latent factors and each other. There are five latent factors or skills in our case, namely literacy ($J_s = 3$), receptive language ($J_s = 2$), math ($J_s = 3$), working memory ($J_s = 2$), non-cognitive-SDQ ($J_s = 4$) and non-cognitive-BPI ($J_s = 6$). All of these latent factors are freely correlated with each other.

Following Anderson and Rubin (1956), we normalize the factor model by setting its scale in such a way that the factor loading on the first of the items of each factor is one; that is, $\lambda_1^s = 1$ for all s. In particular, the normalization measures for each five factors are Thai letter identification, listening comprehension, mental transformation, backward digit span, conduct problems, and headstrong, respectively. See Cunha et al. (2010) for a general identification.

[8] Note that peer-relationship problems from SDQ was dropped because they can not be grouped with the others. An EFA for the following five items, Thai letter identification, English letter identification, word reading, receptive spatial vocabulary and listening comprehension, indicate that the first three are together and the last two form another group. Though, only the eigenvalue of the first group is larger than one. Nevertheless, we form the last two as another group since they are important skills for young children.

The factor model is estimated using a confirmatory factor analysis (CFA) approach (e.g., Gorsuch, 1983; Thompson, 2004) using the full-information maximum likelihood estimation with normally distributed errors.[9] See online appendix D for estimation results.

Fifth, the last step is to predict factor scores of all five latent factors or skills. Following Heckman et al. (2013) and Attanasio et al. (2020), we predict the factor scores for each individual in the data using Bartlett method, which is a generalized least square procedure and leads to an unbiased predictor conditional on factor loadings are known (Bartlett, 1937). Note that a predicted factor score of an observation will be a missing value if one of his/her standardized scores for that particular latent factor is missing.

To assess the informativeness of latent factors, we calculate the signal-to-noise ratio of an item/ measure j for latent factor θ^s based on the following equation:

$$S_j^s = \frac{\left(\lambda_j^s\right)^2 Var\left(\theta^s\right)}{\left(\lambda_j^s\right)^2 Var\left(\theta^s\right) + Var\left(\varepsilon_j^s\right)}, \tag{2}$$

which captures the fraction of the variance of each measure that can be explained by the corresponding latent factor. Technically, this signal-to-noise ratio is equal to one minus its uniqueness, which is equivalent to the R^2 of its measurement equation. The estimation results, reported in table A.1, confirm that all measures are prone to measurement errors as all signal-to-noise ratios are below one. This fact emphasizes the benefit of the factor approach in modeling cognitive and non-cognitive skills. The signal-to-noise ratios for literacy, receptive language, math, working memory, non-cognitive-SDQ and non-cognitive-BPI are between 0.34-0.51, 0.20-0.24, 0.24-0.43, 0.53-0.58, 0.25-0.66 and 0.31-0.61, respectively, which are slightly lower than the ones in Cunha and Heckman (2008), ranged from 0.35 to 0.95 for cognitive skills and from 0.20 to 0.63 for non-cognitive skills.

3.2.2 Treatment Variables: School Closure due to COVID-19 and School Days

The first treatment variable is an indicator for a provincial order to "lockdown" due to the COVID-19 pandemic during January 2021. Since Thailand has adopted a localized lockdown approach, some provinces (28 out of 77 all provinces) were lockdowned and

[9]Under this approach, missing values are assumed to be missing at random and requires that all observed and latent variables are distributed jointly normal.

some were not. This localization feature is clearly different from the national lockdown in Switzerland and Netherlands, utilized in Tomasik et al. (2020) and Engzell et al. (2021), respectively. In our sample, four out of 25 provinces were lockdowned. This variation is key to our identification.

The second treatment variable is an indicator for school closure during the academic year 2020-2021 (from May 2020 to the testing date). This information comes from a telephone interview of teachers.[10] A school will be designated as being closed due to COVID-19 if it answered that it was closed due to COVID-19 for more than 5 school days.[11] See figure 1 for the distribution of the number of days that schools were closed due to COVID-19 for both no-lockdowned and lockdowned provinces. It is evident that schools in lockdowned provinces were closed more than in non-lockdowned provinces. Conditional on being closed at least one day, the average and standard deviation of school-closure days are 14.2 and 7.6 days, respectively, with the maximum of 42 days. In addition, household data reveal that, during school closure, most students received worksheets from schools (about 91%), some had online classes (about 27%), some attended distance learning television program from the central government (about 23%), and few of them did nothing (about 4%).[12] We group all those approaches together as the single alternative to in-person learning.

The third treatment variable - no-school intensity - is the ratio of no-school days and the total days before the survey. No-school days is the number of days children did not go to school from April 1, 2020 (the first day of school break) until the survey date, which include weekends, holidays, school breaks and school closure due to COVID-19.[13] This ratio should have been around 0.37 (135 out of 365 days) if there were no COVID

[10]In fact, we have asked teachers to self-fill the teacher QN during the survey as well, but there is a significant amount of missing data. Therefore, we have revised the QN and interviewed most of them by phone during May and June 2021. We have succeeded in completing the phone interviews for 857 out of 927 schools or 92 percent. For consistency, we dropped samples from schools that could not be interviewed by phone (70 schools).

[11]This threshold is to capture the fact that Kindergarten classes are closed from time to time due to other viruses, e.g., respiratory syncytial virus (RSV), and the closure length is usually about a week. This five-days threshold can be relaxed without significant changes to the estimation results.

[12]Note that each household may report doing several activities during school closure. Therefore, the sum of all fractions is larger than 100%.

[13]This measurement is related with the concept of summer slide (Alexander et al., 2016).

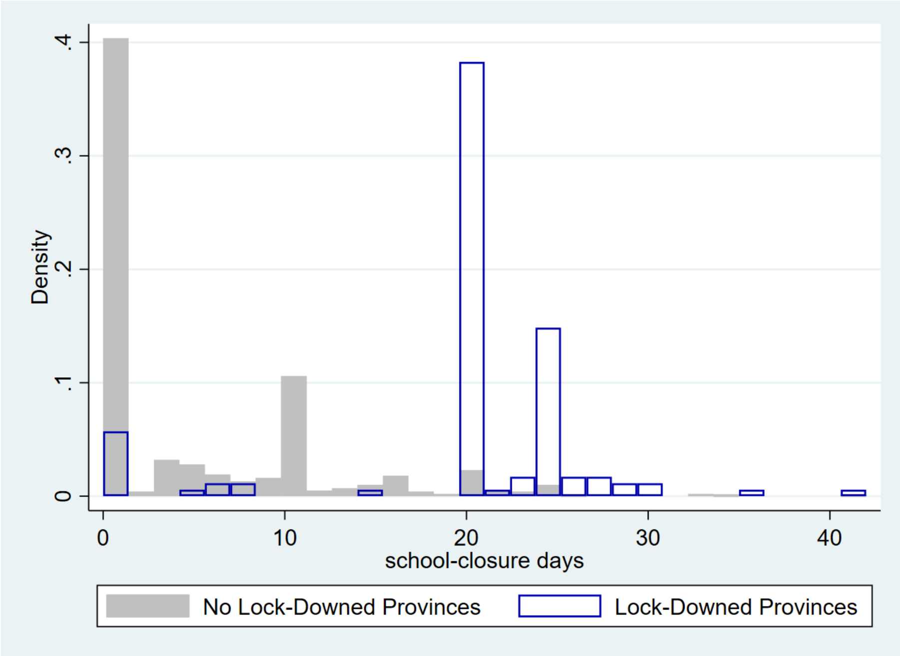


Figure 1: Histograms of school-closure days from the 2021 Thailand school readiness survey for no-lockdowned and lockdowned provinces using school as the unit of observation.

and we were to test on the last day of school. On the other hand, with the outbreak of COVID-19, the average of no-school intensity is about 0.57 with the minimum of 0.44 and the maximum of 0.73. There are three main sources of the variation for this ratio: school closure due to COVID-19, school opening postponement and testing date. See figure 2 for the distribution of no-school intensity.

The last treatment variable is the number of days that children have come to school (face-to-face instruction) between the beginning of the academic year and the testing date, called school days. This variable is to estimate daily learning gains from face-to-face schooling. The difference of school days mainly happened in January and February 2021 (part of the second semester), when the COVID-19 had forced several provinces to go into lockdown. Of course, schools which closed due to COVID tend to have fewer school days than the other group as shown in figure 3. Note that school days can also be varied due to the differences in testing date.

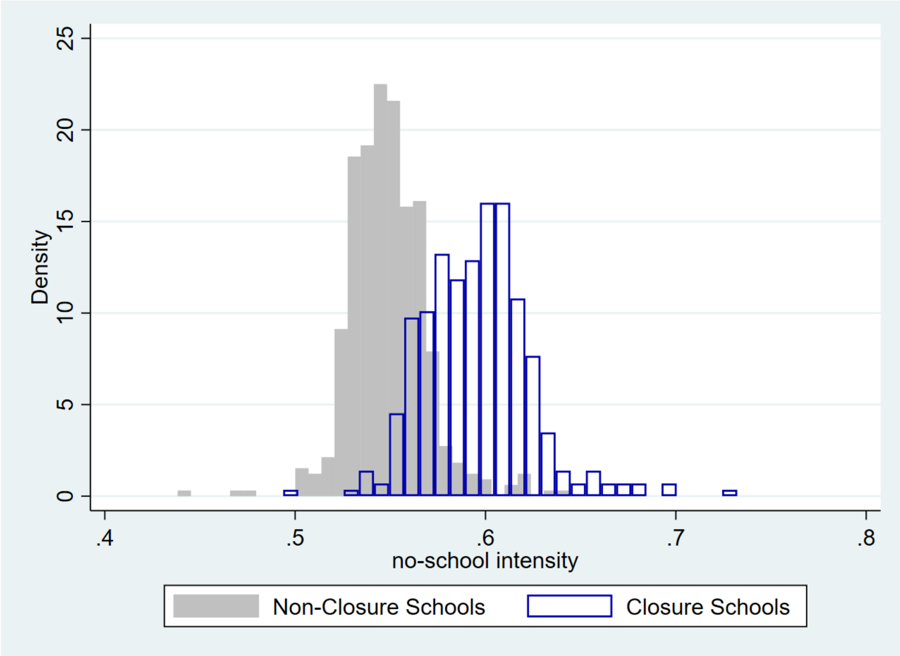


Figure 2: Histograms of no-school intensity from the 2021 Thailand school readiness survey for non-closure and closure schools using school as the unit of observation.

3.3 Estimation Methods

We first estimate the impact of the provincial lockdown order on children skills using the following linear specification:

$$Y^s_{ijp} = \alpha + \beta C_p + \boldsymbol{\gamma} \boldsymbol{X}_{ijp} + \varepsilon^s_{ijp}, \tag{3}$$

where Y^s_{ijp} is the level of skill s of child i attending school j in province p; C_p is an indicator variable for the provincial lockdown whose value is one for a province that was lockdowned due to COVID-19 and zero otherwise; and $\boldsymbol{X}_{ijp}$ are control variables which can be categorized into three groups, macro, school and individual levels. Marco-level variables include the leave-out averages of night light intensity and Grade-6 Ordinary National Educational Test (O-NET) for Thai, English, Mathematics and Sciences at the province and district levels.[14] The first group is to capture the aggregate economic development and the second is to capture the aggregate quality of education. School-level variables include student-teacher ratio, an indicator for private school, an indicator for

[14]For the provincial level leave-out average, the average excludes the sample district while, for the district level, the average excludes the sample school. There are some schools without Grade-6 O-NET scores because they are specialized for kindergarten only. The district-level average for this group is simply the average of the sample district.

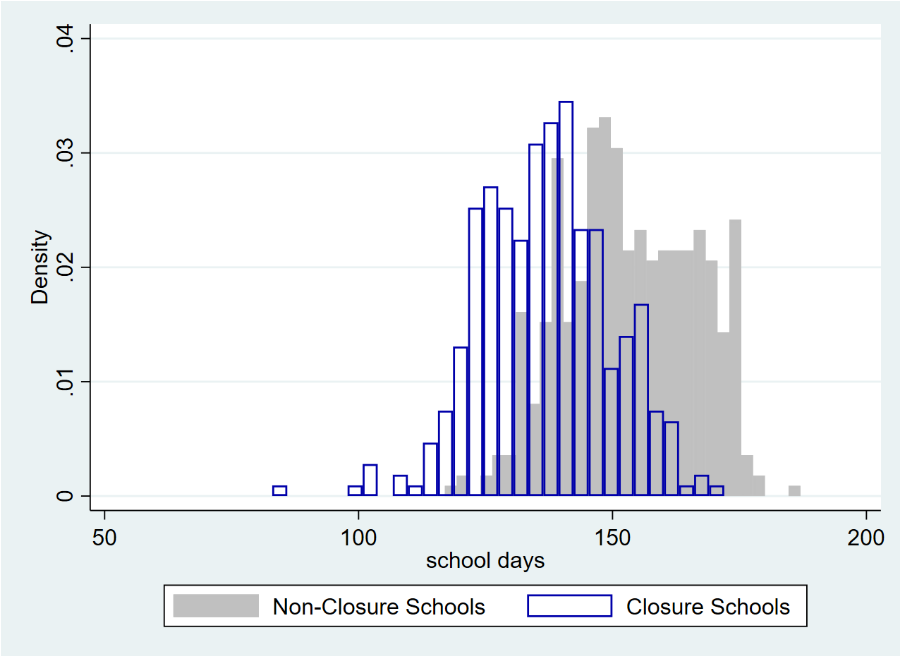


Figure 3: Histograms of school days from the 2021 Thailand school readiness survey for non-closure and closure schools using school as the unit of observation.

being specialized in kindergarten, an indicator of being in urban, fraction of teachers with early childhood education (ECE) degree, fraction of government teachers, fraction of teachers with bachelor degrees, fraction of teachers with master degrees or above. Individual-level variables are child gender, child age, child age squared, child weight, child height, an indicator for special needs (from teacher), an indicator for being sicker more than normal children (from parent), an indicator for using the standard Thai as the main language at home, an indicator for having Thai nationality, household wealth[15], parental absence[16], household size, education level of caregiver (grade 9 which is the current compulsory level or below (reference group); completed grade 12, completed vocational schools, completed bachelor degree or above), caregiver's age, caregiver's age

[15]Household wealth is generated using a confirmatory factor analysis (CFA) method (MLE with one latent factor) and the Bartlett factor score prediction based on the number of the following 8 assets: including cars (sedan), mobile phones, computers, televisions, water heaters, washing machines, air conditioners and refrigerators. This part is different from the CFA performed with the test scores in that here we will drop observations with missing data instead of assuming that missing values are assumed to be missing at random. This is because, for most of the households, the data either contain all asset information or none.

[16]There are 35 percent of children whose main caregivers are neither mother nor father. In fact, there are 19 percent of children where neither mother nor father stays at home with them.

squared, the number of days having breakfast, the number of days parent(s) read to the child, computer and tablet ownership. See table A.2 for summary statistics of key variables.

We estimate this model using the standard ordinary least square (OLS). The identification assumption is that the error terms ε^s_{ijp} are orthogonal to the treatment variable, C_p, conditional of the macro-level control variables. All main results depend on this assumption. We are convinced that this is a reasonable assumption. First of all, the indicator variable for the provincial lockdown, C_p, is driven mainly by the outbreak of COVID, which is a natural phenomenon. Hence, we should be able to consider this shock as a natural experiment. Second of all, the indicator variable for the provincial lockdown, C_p, is a provincial-level variable. Therefore, it should not be directly correlated with individual (e.g., time and material investments) or school variables (e.g., teacher effort). However, as apparent in figure B.1, lockdowned provinces are mostly in the central region, which is relatively wealthier. It is likely that, on average, households and schools in this region invest more in education.[17] To mitigate this concern, the macro-level control variables are included. Even if this indirect correlation exists, it is likely to underestimate the learning loss due to COVID-19, capturing by β, (tend to be more positive than it should be). In addition, we also perform placebo or falsification tests using the school readiness data surveyed in academic year 2020 (prior to the outbreak of COVID-19 in Thailand). The results ensure that our empirical models are valid. See section 7 below.

The second estimation is for the impact of school closure due to the outbreak of COVID-19 using the following linear specification:

$$Y^s_{ijp} = \alpha + \beta C_{jp} + \boldsymbol{\gamma} \boldsymbol{X}_{ijp} + \varepsilon^s_{ijp}, \tag{4}$$

where C_{jp} is an indicator for school closure. First of all, there is a significant amount of non-compliance in school closure. Almost 36 percent of sample schools in non-lockdown provinces were closed (more than 5 days) at least once while about 8 percent in lockdown provinces were not closed. In addition, it is likely that this treatment variable will be

[17] Thailand socio-economic survey (SES) in 2017 indicates that education investment per child in those 28 lockdowned provinces is about 20,452 Baht while the rest is about 9,790 Baht. Averages of grade-6 O-NET scores for Thai, English, Mathematics and Sciences in academic year 2020 for those lockdowned provinces are 50.15, 37.53, 34.10 and 36.69 relative to 45.88, 31.25, 30.57 and 33.59 for the rest of the country.

correlated with the error terms due to omitted variables. For example, some schools in non-lockdown provinces that were closed may have been poorly managed and, therefore, have lower quality. But, teaching quality is unobserved and omitted. To correct for the non-compliance and the endogeneity problem, we apply an instrumental variable approach with the indicator variable for the provincial lockdown, C_p, and the accumulated number of COVID confirmed cases in each province (counting from January 2020 until the test date) as instruments. Note that the argument for the validity of the first instrument is the same as earlier. For the second instrument, we again resort to the fact that the outbreak of COVID is a natural phenomenon.

The same set of instruments are applied to estimate the impact of no-school intensity and the daily learning gains from face-to-face schooling by substituting the treatment variable, C_{jp}, by no-school intensity and school days, respectively.

In addition, we estimate heterogeneous effects for household characteristics (parental absence, wealth, household size, parental education), children characteristics (child gender, child age, having special needs), parental investments (having tutoring, number of days read to child, the number of information technology gadgets per head, the number of computers and tablets per head), and school characteristics (fraction of teachers with ECE degree). These heterogeneity variables are considered each one at a time. Note that the heterogeneous effect for a variable of interest is measured by the estimation coefficient of the interaction term between the variable and the corresponding treatment variable, which is the school closure dummy.

Technically, to account for the family-wise error rate (FWE), which is the probability of falsely rejecting at least one true null hypothesis, or type I errors when performing multiple hypotheses tests, we apply the step-down approach of Romano and Wolf (2005, 2016) on each of the two sets of outcomes, cognitive and non-cognitive skills, with 1,000 bootstrap replications and classroom level clustering, and report the associated p-values for each set of outcomes.[18] Symbols *, ** and *** denote 0.10, 0.05 and 0.01 significant level with the Romano and Wolf correction.

[18]This correction is done using a STATA command: rwolf2, which created by Clarke (2021).

4 Main Results

This section presents empirical evidence indicating that school closure during the outbreak of COVID-19 has caused a negative impact on kindergartners' cognitive skills significantly but not for non-cognitive skills. Daily learning gains are also presented in this section. All related first-stage regression and ordinary least square results are reported in the online appendix E and F.

Panel A of table 1 shows the impact of the provincial lockdown order on children skills. Most of the estimation coefficients are negative except for receptive language and non-cognitive-BPI (from parents). However, only the estimation coefficient for working memory is negative and statistically significant with a p-value with Romano-Wolf correction of 0.001. More specifically, the effect of the provincial lockdown order on working memory is about 0.321 standard deviation (SD), which has practical significance. In other words, young children in lockdowned provinces performed significantly worse than the others especially in the working memory test.

The results in panel B of table 1 are even stronger. The estimation coefficients for school closure dummy are all negative except for non-cognitive-BPI (from parents). The estimation coefficients are negative and significant not only for working memory but also math. More specifically, closing schools due to COVID-19 had reduced children scores for math and working memory by 0.452 ($p = 0.017$) and 1.677 ($p = 0.002$) SD, respectively, all of which have practical significance. These results imply that school closure during the outbreak has caused substantial learning losses for Thai kindergartners, especially in cognitive domains.

Technically, the instruments are relevant and strong enough to avoid weak instrument bias. First, Kleibergen-Paap F-statistics (Kleibergen and Paap, 2006) for the school closure estimations are well above 10, the popular threshold proposed by Stock et al. (2002), and, correspondingly, their p-values are smaller than 0.05. Second, p-values for underidentification tests (rank tests) are virtually zero. Third, we cannot reject overidentification tests, using Hansen-J statistics, for all but receptive and math. This result suggests that our instruments are orthogonal to the error terms in most cases.

Similarly, the estimation coefficients for no-school intensity are all negative except for non-cognitive-BPI (from parents). See panel A of table 2. The estimation coefficients are negative and significant for both math and working memory with the effect size about

Table 1: Estimation results for the impact of provincial lock-down and school closure.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: Provincial lock-down						
impact	-0.0711	0.0672	0.00529	-0.321***	0.0197	0.0282
RW p-value	[0.349]	[0.349]	[0.879]	[0.001]	[0.756]	[0.738]
p-value	(0.193)	(0.134)	(0.880)	(0.000)	(0.757)	(0.474)
No. Obs.	7805	7805	7805	6114	7503	7804
Panel B: School closure						
impact	-0.210	-0.268	-0.452**	-1.677***	-0.0932	0.163
RW p-value	[0.249]	[0.249]	[0.017]	[0.002]	[0.626]	[0.357]
p-value	(0.205)	(0.103)	(0.001)	(0.000)	(0.632)	(0.188)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	37.44	37.44	37.44	27.06	36.86	37.51
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.495	0.001	0.000	0.693	0.461	0.638

Note: RW p-values and p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications and the traditional p-values both clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * p<0.10, ** p<0.05, *** p<0.01.

3.949 (p = 0.014) and 16.31 (p = 0.001) SD, respectively. Intuitively, an additional day[19] of no-school causes learning losses in math and working memory by 0.0116 ($\frac{3.949}{339}$) and 0.0481 ($\frac{16.31}{339}$) SD, respectively. Again, these results indicate that school closure during the outbreak has caused significant learning losses for Thai kindergartners, especially in cognitive domains. The instruments are shown to be valid as in the case of school closure. See all test statistics in panel A.

Table 2: Estimation results for the impact of no-school intensity and logarithm of no-school intensity.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: No-school intensity						
impact	-2.333	-1.943	-3.949**	-16.31***	-0.734	1.579
RW p-value	[0.355]	[0.355]	[0.014]	[0.001]	[0.713]	[0.419]
p-value	(0.174)	(0.231)	(0.002)	(0.000)	(0.714)	(0.226)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	42.37	42.37	42.37	35.83	41.37	42.39
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.595	0.000	0.000	0.500	0.432	0.508
Panel B: Log of no-school intensity						
impact	-1.354	-1.148	-2.313***	-9.501***	-0.432	0.921
RW p-value	[0.351]	[0.351]	[0.004]	[0.001]	[0.722]	[0.406]
p-value	(0.174)	(0.225)	(0.002)	(0.000)	(0.711)	(0.224)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	45.79	45.79	45.79	38.89	44.71	45.83
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.590	0.000	0.000	0.528	0.433	0.512

Note: RW p-values and p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications and the traditional p-values both clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * p<0.10, ** p<0.05, *** p<0.01.

Table 2 also shows the estimation results using the logarithm of no-school intensity

[19]This calculation is based on the average total days before test, which is 339 days. The average comes from the data set of 7,489 sample, where all missing data are dropped.

instead of the ratio, as a robustness check. This specification allows us to interpret the estimation in terms of percentage change. See panel B of the table. The results are similar to the results in panel A. The estimation coefficients are negative and significant for both math and working memory with the effect size about 2.313 ($p = 0.004$) and 9.501 ($p = 0.001$) SD, respectively. Roughly speaking, an additional day[20] of no-school causes learning losses in math and working memory by 0.0120 ($\frac{2.313}{193}$) and 0.0492 ($\frac{9.501}{193}$) SD, respectively, which are closed to the corresponding results above. Again, the instruments are shown to be valid as in the case of school closure. See all test statistics in panel B.

Table 3 presents daily learning gains from having face-to-face instruction at school, which is the opposite side of the story. The results indicate that going to school has a significant benefit to Thai kindergartners, especially in receptive language, math and working memory. The estimation coefficients for school days are all positive except for non-cognitive-BPI (from parents). The estimation coefficients are positive and significant for receptive language, math and working memory with the effect size about 0.0094 ($p = 0.025$), 0.0130 ($p = 0.001$) and 0.0341 ($p = 0.001$) SD, respectively. In other words, an additional day of schooling produces learning gains in receptive language, math and working memory by 0.0094, 0.0130, and 0.0341 SD, respectively. Again, these results should serve as strong evidence showing that going to school has a meaningful benefit to kindergartners. The instruments are shown to be valid as in the case of school closure. See all test statistics in panel A.

Panel B of the table also shows the estimation results using the logarithm of school days instead of the school days, as a robustness check. This specification allows us to interpret the estimation in terms of percentage change. See panel B of table 3. The results are similar to the results in panel A. The estimation coefficients are positive and significant for receptive language, math and working memory with the effect size about 1.334 ($p = 0.024$), 1.838 ($p = 0.001$) and 4.780 ($p = 0.001$) SD, respectively. Note that the treatment variable here is the logarithm of school days, and, therefore, we should interpret these effects as percentage changes. Roughly speaking, an additional day of schooling[21] produces learning gains in receptive language, math and working memory by

[20]This calculation is based on the average no-school days, which is 193 days. The average comes from the data set of 7,489 sample, where all missing data are dropped. Note that adding an additional day of no-school will lead to an increase in no-school intensity by one divided the average no-school days, which is $\frac{1}{193}$ percent in this case.

[21]This calculation is based on the average school days, which is 146 days. The average comes from the

Table 3: Estimation results for the impact of school days and logarithm of school days.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: School days						
impact	0.0040	0.0094**	0.0130***	0.0341***	0.0032	-0.0040
RW p-value	[0.265]	[0.025]	[0.001]	[0.001]	[0.454]	[0.286]
p-value	(0.266)	(0.011)	(0.000)	(0.000)	(0.483)	(0.146)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	46.56	46.56	46.56	40.91	44.72	46.58
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.346	0.005	0.003	0.0990	0.548	0.889
Panel B: Log of school days						
impact	0.557	1.334**	1.838***	4.780***	0.452	-0.568
RW p-value	[0.282]	[0.024]	[0.001]	[0.001]	[0.477]	[0.258]
p-value	(0.272)	(0.010)	(0.000)	(0.000)	(0.478)	(0.146)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	40.12	40.12	40.12	35.17	38.49	40.14
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.341	0.006	0.007	0.0943	0.552	0.901

Note: RW p-values and p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications and the traditional p-values both clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * p<0.10, ** p<0.05, *** p<0.01.

0.0091 ($\frac{1.334}{146}$), 0.0126 ($\frac{1.838}{146}$) and 0.0327 ($\frac{4.780}{146}$) SD, respectively, which are closed to the corresponding results above. Again, the instruments are shown to be valid as in the case of school closure. See all test statistics in panel B.

In addition, we perform a back-of-envelope calculation, based on the estimation results for school closure, no-school intensity and school days, to measure learning losses using school days as a unit. First, closing schools due to the outbreak of COVID-19, on average, causes learning losses in math and working memory by 34.8 ($\frac{0.452}{0.0130}$) and 49.2 ($\frac{1.677}{0.0341}$) school days, respectively, while the maximum number of days of school closure was 42 school days. Second, an additional day of no-school causes learning losses in math and working memory by 0.92 ($\frac{0.0120}{0.0130}$) and 1.44 ($\frac{0.0492}{0.0341}$) school days, respectively. To sum up, there are significant losses when young children could not go to school.

5 Heterogeneous Effects of School Closure

This section answers whether school closure due to the outbreak of COVID-19 affects young children differently across subgroups, including household characteristics (parental absence, wealth, household size, parental education), children characteristics (child gender, child age, having special needs), parental investments (having academic tutoring, number of days read to child, the number of information technology gadgets per head, the number of computers and tablets per head), and school characteristics (fraction of teachers with ECE degree). To save space, we present the estimation results for school closure dummy only. Technically, we apply an instrumental variable approach using a provincial lockdown indicator and the number of COVID confirmed cases, and their interactions with the variable of interest as instruments.

Panel A of table 4 shows the heterogeneous impact of school closure on children skills with respect to child gender. The result indicates that female students have been affected by school closure significantly more than male, especially in working memory. See the "sch close x female" row in panel A. This does not mean that female students have lower working memory, on average. On the contrary, they still perform significantly better in all but receptive language. See the "female" row in panel A where most of the coefficients

data set of 7,489 sample, where all missing data are dropped. Note that an additional day of schooling will lead to an increase in school days by one divided the average school days, which is $\frac{1}{146}$ percent in this case.

are positive. Intuitively, these results imply that females may benefit from going to school more than males, and school closure took away that advantage from them. Therefore, they have been more adversely affected by school closure.

Panel B of table 4 shows the heterogeneous impact of school closure with respect to an indicator for special needs (reported by teacher). The result indicates that special-needs students have been affected by school closure significantly less than the other group. See the “sch close x needs” row in panel A showing that all estimation coefficients are positive, and statistically significant for literacy and non-cognitive skills (SDQ). Of course, they still perform significantly lower in all domains. See all negative and significant coefficients in the “needs” row in panel B. Intuitively, these results imply that special-needs students may need personalized and intensive care, which is difficult to receive in average schools in Thailand. On the other hand, school closure means parents have to personally provide the care, which is supposed to be more intensive, and, therefore, benefit special-needs students significantly. However, this may come with a sizable cost, e.g., a parent may have to leave the labor force. Unfortunately, we have no information to estimate such cost at this point.

The heterogeneous impact of school closure with respect to household wealth is presented in panel A of table 5. The result indicates that wealthier children have been affected by school closure significantly more, especially in math. See the “sch close x wealth” row in panel A showing that estimation coefficients are negative for all but non-cognitive (BPI), and statistically significant for math. This result may seem surprising at first because wealthier families should have more resources (e.g., computers, tablets) to facilitate remote learning. But it will make more sense when we consider together with the heterogeneous impact of the school closure with respect to having academic tutoring (see panel B of table 5). In addition, as will be shown below, having computers or tablets does not seem to help mitigate the impact (see table 7). Again, this does not mean that wealthier children perform worse than the others. In fact, they still performed much better, as shown in the “wealth” row in panel A.

The heterogeneous impact of school closure with respect to private academic tutoring is shown in panel B of table 5. The result indicates that children who have private academic tutoring were affected by school closure significantly more, especially in math. See the “sch close x tutoring” row in panel B showing that estimation coefficients are all

Table 4: Heterogeneous effects with respect to child gender and special needs.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: Interaction with an Indicator for Female						
school closure	-0.149	-0.289	-0.459**	-1.490***	-0.0966	0.163
RW p-value	[0.382]	[0.194]	[0.011]	[0.004]	[0.649]	[0.387]
p-value	(0.367)	(0.081)	(0.000)	(0.000)	(0.624)	(0.203)
female	0.167***	-0.0956*	0.0621*	0.237***	0.320***	0.119***
RW p-value	[0.001]	[0.054]	[0.081]	[0.002]	[0.001]	[0.008]
p-value	(0.000)	(0.022)	(0.068)	(0.000)	(0.000)	(0.010)
sch close x female	-0.118	0.0753	0.0262	-0.374***	0.00719	-0.0178
RW p-value	[0.284]	[0.567]	[0.731]	[0.008]	[0.964]	[0.964]
p-value	(0.100)	(0.347)	(0.693)	(0.001)	(0.940)	(0.835)
F-statistics	21.86	21.86	21.86	17.07	21.29	21.87
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.769	0.000	0.001	0.847	0.759	0.526
Panel B: Interaction with an Indicator for Special Needs (from teacher)						
school closure	-0.232	-0.279	-0.468***	-1.661***	-0.139	0.175
RW p-value	[0.211]	[0.211]	[0.006]	[0.003]	[0.471]	[0.299]
p-value	(0.159)	(0.092)	(0.000)	(0.000)	(0.474)	(0.160)
needs	-0.503***	-0.556***	-0.668***	-0.349**	-1.490***	-0.366***
RW p-value	[0.001]	[0.001]	[0.001]	[0.030]	[0.001]	[0.002]
p-value	(0.000)	(0.000)	(0.000)	(0.040)	(0.000)	(0.000)
sch close x needs	0.394*	0.0957	0.262	0.0109	0.595**	0.0112
RW p-value	[0.066]	[0.898]	[0.315]	[0.971]	[0.013]	[0.952]
p-value	(0.009)	(0.676)	(0.116)	(0.972)	(0.003)	(0.951)
F-statistics	20.14	20.14	20.14	14.40	19.52	20.22
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.728	0.004	0.001	0.273	0.718	0.337

Note: RW p-values and p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications and the traditional p-values both clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 5: Heterogeneous effects with respect to household wealth and having tutoring lessons.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: Interaction with Household Wealth Index						
school closure	-0.183	-0.253	-0.439**	-1.617***	-0.0944	0.162
RW p-value	[0.264]	[0.222]	[0.014]	[0.003]	[0.632]	[0.368]
p-value	(0.262)	(0.117)	(0.001)	(0.000)	(0.622)	(0.191)
wealth	0.132**	0.0452	0.0988**	0.195*	0.0549	-0.00592
RW p-value	[0.017]	[0.184]	[0.012]	[0.072]	[0.361]	[0.871]
p-value	(0.001)	(0.189)	(0.001)	(0.019)	(0.176)	(0.860)
sch close x wealth	-0.0908	-0.0534	-0.105*	-0.260	-0.0425	0.0169
RW p-value	[0.231]	[0.350]	[0.088]	[0.144]	[0.725]	[0.725]
p-value	(0.118)	(0.331)	(0.011)	(0.031)	(0.470)	(0.737)
F-statistics	18.66	18.66	18.66	12.84	18.64	18.69
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.198	0.001	0.001	0.738	0.418	0.868
Panel B: Interaction with an Indicator for Having Tutoring						
school closure	-0.134	-0.209	-0.410**	-1.594***	-0.00524	0.186
RW p-value	[0.414]	[0.412]	[0.017]	[0.002]	[0.983]	[0.264]
p-value	(0.408)	(0.209)	(0.002)	(0.000)	(0.978)	(0.136)
tutoring	0.199**	0.122	0.129**	0.115	-0.0176	-0.0374
RW p-value	[0.014]	[0.115]	[0.045]	[0.335]	[0.831]	[0.820]
p-value	(0.002)	(0.059)	(0.015)	(0.351)	(0.828)	(0.590)
sch close x tutoring	-0.239	-0.243	-0.254**	-0.358	-0.0360	-0.0425
RW p-value	[0.131]	[0.131]	[0.046]	[0.163]	[0.932]	[0.932]
p-value	(0.036)	(0.033)	(0.007)	(0.128)	(0.806)	(0.734)
F-statistics	18.30	18.30	18.30	12.90	18.05	18.33
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.570	0.003	0.002	0.805	0.728	0.635

Note: RW p-values and p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications and the traditional p-values both clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * p<0.10, ** p<0.05, *** p<0.01.

negative, and statistically significant for math. This result helps explain why wealthier children surprisingly have been negatively affected more by school closure. In normal time, private tutoring or summer camps are more accessible and beneficial to wealthier children, as discussed in the summer slide literature (e.g., Alexander et al., 2007). However, the outbreak of COVID-19 disallows such activities and that took away the advantage from the wealthier children. As a result, they have larger learning losses relative to what could have been without COVID. Note that private academic tutoring still has positive and significant impact on literacy and math, as shown in the "tutoring" row in panel B.

Another surprising result is the heterogeneous impact of school closure with respect to caregiver's education. The result in panel A of table 6 indicates that children whose caregiver finished college or above have been affected by school closure significantly more, especially in non-cognitive skills (SDQ). See the "sch close x college" row in panel A showing that estimation coefficients are negative for all but literacy, and statistically significant for non-cognitive skills (SDQ). Intuitively, this can be explained using the same argument as in the case of wealth. That is, better educated caregivers are more likely to hold full time jobs with higher income and, therefore, resort to schools or private academic tutoring to compensate for their scarcity of time. Unfortunately, the outbreak of COVID-19 took away those options and, therefore, their children have been adversely affected. Again, this does not mean that children with better educated caregivers perform worse than the others. In fact, they still performed much better, as shown in the "college" row in panel A.

Panel B of table 6 shows the heterogeneous impact of school closure with respect to parental absence. The result indicates that children whose biological parents are not home have been affected by school closure significantly more than the other group, especially in non-cognitive skills (both SDQ and BPI). See the "sch close x absence" row in panel A showing that estimation coefficients are negative for all but working memory, and statistically significant for non-cognitive skills. This result is as expected and worrisome. Children with parental absence usually live with grandparents, who may have difficulty to facilitate learning at home. This result implies that children with parental absence rely on schools more than the others. On the other hand, absent Thai parents have to be away from home mostly to find jobs, and they usually send back significant amount of

Table 6: Heterogeneous effects with respect to main caregiver education and parental absence.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: Interaction with an Indicator for Caregiver with College or Above						
school closure	-0.202	-0.224	-0.443***	-1.649***	-0.0582	0.181
RW p-value	[0.365]	[0.365]	[0.009]	[0.003]	[0.761]	[0.309]
p-value	(0.220)	(0.169)	(0.001)	(0.000)	(0.762)	(0.148)
college	0.342***	0.333***	0.159**	0.164	0.222**	0.0750
RW p-value	[0.001]	[0.001]	[0.021]	[0.198]	[0.019]	[0.398]
p-value	(0.000)	(0.000)	(0.015)	(0.192)	(0.008)	(0.380)
sch close x college	0.0320	-0.323	-0.0386	-0.0233	-0.350*	-0.0627
RW p-value	[0.973]	[0.203]	[0.973]	[0.973]	[0.060]	[0.692]
p-value	(0.840)	(0.057)	(0.741)	(0.911)	(0.016)	(0.678)
F-statistics	18.83	18.83	18.83	13.62	18.43	18.86
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.322	0.004	0.001	0.467	0.518	0.245
Panel B: Interaction with an Indicator for Parental Absence						
school closure	-0.188	-0.234	-0.430**	-1.683***	-0.0338	0.241
RW p-value	[0.328]	[0.328]	[0.015]	[0.002]	[0.862]	[0.111]
p-value	(0.260)	(0.150)	(0.001)	(0.000)	(0.863)	(0.060)
parental absence	0.0394	0.100	0.0167	-0.110	0.0901	0.162**
RW p-value	[0.624]	[0.225]	[0.724]	[0.438]	[0.138]	[0.044]
p-value	(0.407)	(0.068)	(0.709)	(0.196)	(0.148)	(0.019)
sch close x absence	-0.116	-0.169	-0.107	0.0467	-0.286**	-0.392**
RW p-value	[0.477]	[0.310]	[0.477]	[0.757]	[0.026]	[0.014]
p-value	(0.198)	(0.088)	(0.203)	(0.756)	(0.016)	(0.003)
F-statistics	18.54	18.54	18.54	13.57	18.29	18.57
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.110	0.002	0.001	0.721	0.447	0.846

Note: RW p-values and p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications and the traditional p-values both clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * p<0.10, ** p<0.05, *** p<0.01.

Table 7: Heterogeneous effects with respect to computer or tablet ownership, number of days read to child, fraction of teachers with ECE degree, child age, household size and private school.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: Interaction with the number of IT gadgets per head						
sch close x comp	-0.0900	0.000920	-0.0718	-0.193	-0.0720	0.0807
RW p-value	[0.811]	[0.995]	[0.811]	[0.772]	[0.740]	[0.740]
p-value	(0.417)	(0.994)	(0.421)	(0.308)	(0.538)	(0.476)
Panel B: Interaction with the number of computers or tablets per head						
sch close x comp	0.0440	-0.103	0.0482	-0.174	0.0347	0.337
RW p-value	[0.988]	[0.988]	[0.988]	[0.988]	[0.918]	[0.376]
p-value	(0.860)	(0.719)	(0.805)	(0.635)	(0.889)	(0.189)
Panel C: Interaction with number of days read to child						
sch close x read	0.0121	-0.0454	-0.00378	-0.0393	0.0248	-0.00460
RW p-value	[0.819]	[0.443]	[0.845]	[0.632]	[0.524]	[0.819]
p-value	(0.590)	(0.129)	(0.840)	(0.280)	(0.327)	(0.841)
Panel D: Interaction with fraction of teachers with ECE degree						
sch close x ece	0.00924	0.332	0.373	0.313	-0.118	0.0613
RW p-value	[0.954]	[0.238]	[0.178]	[0.665]	[0.834]	[0.834]
p-value	(0.956)	(0.067)	(0.034)	(0.407)	(0.566)	(0.640)
Panel E: Interaction with child age						
sch close x age	-0.0245	-0.153	-0.00868	0.0852	0.169	0.0960
RW p-value	[0.971]	[0.718]	[0.971]	[0.971]	[0.396]	[0.474]
p-value	(0.837)	(0.292)	(0.936)	(0.714)	(0.234)	(0.474)
Panel F: Interaction with household size						
sch close x size	-0.0107	0.0188	-0.0159	-0.0192	0.0114	-0.00506
RW p-value	[0.817]	[0.817]	[0.817]	[0.817]	[0.867]	[0.867]
p-value	(0.528)	(0.381)	(0.363)	(0.563)	(0.627)	(0.806)
Panel G: Interaction with private school						
sch close x private	0.847	-0.230	-0.190	0.155	-1.525	-0.133
RW p-value	[0.565]	[0.918]	[0.918]	[0.918]	[0.481]	[0.714]
p-value	(0.098)	(0.589)	(0.513)	(0.801)	(0.156)	(0.687)

Note: RW p-values and p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications and the traditional p-values both clustered at classroom level. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * p<0.10, ** p<0.05, *** p<0.01.

remittance[22], which, in turn, allows their children to have more resources than otherwise. That is consistent with the fact that all coefficients for parental absence are positive except for working memory. See the "absence" row in panel A. Nevertheless, the outbreak of COVID-19 has shown that this group of children is vulnerable.

Table 7 presents all heterogeneous effects with respect to the other variables including the number of information technology gadgets (computers, tablets and mobile phones) per head, the number of computers and tablets per head, the number of days read to child, fraction of teachers with ECE degree, child age, and household size, all of which are not statistically significant. To save space, we present only the estimation results of the interaction terms and discuss only the few interesting issues. The first issue is the number of information technology gadgets per head, which does not seem to help much. Most of the estimation coefficients are negative though not significant. The results are slightly different for the case of the number of computers and tablets per head, in which the estimation coefficients are now positive but all are insignificant still. See panel A and B of the table. This insignificance may result from the fact that remote learning for kindergartners is not effective to begin with, or, alternatively, they did not participate in the process due to lack of internet access. Unfortunately, we cannot distinguish these two mechanisms since we do not have data on internet access. In addition, we do not know if this finding can be generalized to older age groups since it might be the case that remote learning is more effective for older students. Another issue is reading time, which again does not seem to be as promising as we hoped for. See panel C of the table. It could be that parents did not or could not increase reading time to compensate for school closure due to their scarcity of time.

6 Robustness

This section presents estimation results with changes in several dimensions for robustness checks. The overall results confirm the benchmark results indicating that the outbreak of COVID-19 causes enormous learning losses in cognitive skills for Thai kindergartners. For exposition purposes, the associated tables are shown in appendix A.

[22]The data from rural Thailand in Dinh and Kilenthong (2021) indicates that, for the parental absence families, remittance was about 71 percent of household income.

6.1 Estimation Results with an Additional Mathematics Domain

The subsection alters the benchmark model by incorporating an additional domain in mathematics, namely symbolic addition. This domain was originally excluded because it contained a number of missing values, which result from the assessment design under which only children who could identify one-digit numbers perfectly can be tested. In order to incorporate this domain, we still use the full-information maximum likelihood with normally distributed errors but now predict factor scores using the STATA-default method, the regression-based method, not the Bartlett. This approach estimates factor scores for all observations including the ones with missing values. As a result, the number of observation is now the same for all outcomes.

Table A.3 confirms that our benchmark results are robust. The negative impact on cognitive skills are significant in all domains for most cases while the impact on non-cognitive skills are still insignificant. Note that the effect sizes are smaller than the benchmark models in most cases. Nevertheless, a back-of-envelope calculation gives a very similar picture. More specifically, closing school due to the outbreak of COVID-19, on average, causes learning losses in math and working memory by 36.9 and 42.2 school days, which are comparable to the benchmark results discussed at the end of section 4.

6.2 Estimation Results when the Outcomes are derived using an Alternative Approach

The subsection changes the calculation method for the outcome variables. More specifically, raw scores for literacy, receptive language and math are derived using item response theory (IRT), based on 15, 9 and 13 items, respectively, while a raw score for working memory is the sum of forward and backward scores. The final outcomes for these domains are the age-standardized scores derived using Kernel-weighted Local Polynomial Smoothing up to the third degree polynomial. On the other hand, the outcomes for SDQ and BPI are derived using the same method as in the benchmark case but with no correlation allowed.

Table A.4 again confirms that our benchmark results are robust. The results are qualitatively the same as the benchmark case but quantitatively slightly different. A back-of-envelope calculation gives a similar picture. More specifically, closing schools due to the outbreak of COVID-19, on average, causes learning losses in math and working memory by 33.9 and 49.1 school days.

6.3 Estimation Results with Weighting

The subsection presents estimation results when sampling weights are applied. The sampling weights are calculated based on the stratification method explained in section 3, response rates of schools in each province and the total number of students in each province.[23] See the online appendix G for the sampling weight calculation.

Table A.5 suggests that our benchmark results are robust. Though, the results are only significant for working memory domain for all cases while math domain is now significant only for the effect of school days. This may not be surprising since weighting can potentially reduce the estimation efficiency, as discussed in Solon et al. (2015). Nevertheless, a back-of-envelope calculation gives a similar picture. That is, closing schools due to the outbreak of COVID-19, on average, causes learning losses in working memory by 43.9 school days.

6.4 Estimation Results with the Original Sample

This subsection responds to concerns regarding sample attrition due to missing data. Recall that the original sample contains 11,478 children but the benchmark sample has 7,805 children or 68 percent of the original, due to missing data. Unfortunately, this attrition leads to significant differences between the original and baseline samples as presented in table A.2. Therefore, it is important to check if the results are robust with respect to this dimension.

To keep all the original sample in the analysis, we have to drop most control variables. More specifically, the control variables now have only 10 variables, including the leave-out averages of Grade-6 O-NET and the night light intensity at the province and district levels, child gender, child age, child age squared, child weight, child height, and an indicator for private school.

Table A.6 again confirms that our benchmark results are robust. The results are even stronger with respect to statistical significance. The negative effects are now significant

[23]The total number of students are calculated from the administrative database of the Ministry of Education of Thailand. The authors are grateful for their collaboration.

for not only math and working memory but also literacy. Quantitatively, these results are more similar to the benchmark case. A back-of-envelope calculation gives a similar picture. More specifically, closing schools due to the outbreak of COVID-19, on average, causes learning losses in math and working memory by 29.6, and 36.7 school days.

7 Placebo Tests

This section presents placebo or falsification tests using test scores from the 2020 Thailand school readiness data, collected during January and March 2020, and information on school closure of the same schools surveyed in 2021. The key point is that this data contain the same set of outcome and control variables[24] and were surveyed before the COVID-19 outbreak in Thailand. It also has a weakness. That is, this survey covered 19 provinces, which are different from the 25 provinces in our main sample. Nevertheless, four of 19 provinces were later lockdowned due to COVID-19 outbreak in 2021. See figure B.3 for sampled provinces in this 2020 survey.

In order to obtain information regarding school closure in 2021 for the 2020 survey sample, we interviewed teachers in the 2020 sample by phone in July 2021 using the same questionnaire as in the 2021 survey and applied the same calculation procedure described in section 3.2.2.[25] School-closure days, no-school intensity and school days used in the placebo tests are comparable to the main sample. Compare figure 1 with B.4, figure 2 with B.5, and figure 3 with B.6, shown in appendix B. This additional information allows us to perform the tests based on all four empirical models. In addition, with a rich set of information, we can estimate the model with the same set of control variables.

The main hypotheses for our placebo tests are that provincial lockdown order, school closure and closure intensity should not cause an adverse on children skills since the data were collected one year before the COVID-19 outbreak in Thailand. That is, we would feel more confident with the benchmark results if our placebo tests could not reject the null hypothesis that there is no adverse effect on children skills.

Technically, we again account for the family-wise error rate (FWE) when performing

[24]To be more precise, there are small differences in terms of the outcomes. That is, non-cognitive skills from the teacher QN are derived from the BPI questions instead of the SDQ. See table A.8 for summary statistics of key variables.

[25]We have succeeded in completing the phone interviews for 626 out of 684 schools or 92 percent, which is almost the same rate as in the 2021 survey.

multiple hypotheses tests using the step-down approach of Romano and Wolf (2005, 2016) on each of the two sets of outcomes, cognitive and non-cognitive skills, with 1,000 bootstrap replications and classroom level clustering.

Table A.7, in appendix A, presents the estimation coefficients of the treatment variables for each outcomes.[26] The overall results confirm that provincial lockdown order, school closure and closure intensity in 2021 do not cause any negative and significant effect on children skills measured in 2020. More specifically, the placebo effects of provincial lockdown are all positive except for non-cognitive reported by teachers. The effects for literacy and receptive language are significant but positive, nonetheless. This suggests that there may be systematic variation across these provinces. This problem would lead to an underestimation of learning losses, however. Similar results are for the impact of school closure, presented in panel B. The results in panel C and D are even more striking. None of the estimation coefficients is statistically significant at 0.10. To sum up, the test results suggest that we should not be concerned that unobserved heterogeneity may contaminate our results (Jack and Suri, 2014).

8 Conclusion and Discussion

This paper shows that school closure during the outbreak of COVID-19 caused enormous learning losses in cognitive skills for Thai kindergartners, especially in mathematics and working memory. Our back-of-envelope calculation indicates that school closure, on average, leads to learning losses equivalent to 35 and 49 school days for math and working memory, respectively, while the maximum number of days of school closure was 42 school days. Roughly speaking, learning losses in cognitive skills for kindergartners was at least 83% for math and more than 100% for working memory. Fortunately, we found no adverse impact on non-cognitive skills, but one should bear in mind that they were derived from self-reporting by teachers and parents, which may contain considerable measurement errors relative to cognitive skills from direct assessments. In addition, our benchmark results confirm that going to school has significantly benefited young children, especially in receptive language, mathematics and working memory.

Another key message from this paper is that the outbreak of COVID-19 has affected

[26] P-values are for the two-sided test with a Romano-Wolf correction (Romano and Wolf, 2005, 2016), whose null hypothesis is that the estimation coefficient of a variable of interest is zero.

heterogeneously across groups, especially children who otherwise might have been more advantageous. This is different from the summer slide story in that schooling interruption due to COVID-19 has put a physical barrier to not just formal schooling but also other means, e.g., private academic tutoring. Wealthier children could not escape this adverse effect because young children still learn through human interaction and stimulation, which are not easy to achieve through online learning. This does not mean that school closure due to COVID-19 did not affect disadvantaged children. On the other hand, wealthier children should be able to recover learning losses quicker once the barrier is lifted. As shown in table A.9, an additional school day is more productive for wealthier children and children with private academic tutoring as all estimation coefficients for the interaction terms are all positive (but not significant with respect to Romano-Wolf correction). This finding also implies that the school day should be interpreted broadly including a possibility of having private academic tutoring. An optimal public policy, therefore, should focus on how to help disadvantaged children recovering their losses.

This paper should contribute to the policy debate regarding school closure since it provides concrete evidence for the cost of school closure. Of course, policy makers need to weigh between its benefit (as to lower the risk of infection) and its cost (learning losses). Another relevant policy question is what to do after the pandemic is under control. A possible policy, for the next couple of years, is to keep schools open during the summer breaks or organize summer schools. This straightforward and simple idea follows naturally from the evidence that their current schools can promote school readiness significantly. Of course, improving school quality should be our long term goal as always.

One concern is that we do not observe prior skills of children and that could lead to an overestimation of the impact if prior skills are positively correlated with school closure. The only way we can mitigate this concern indirectly is to resort to the placebo tests, which indicate that unobserved heterogeneity problems in our benchmark analysis should be negligible. Another weakness is the lack of parental investment information, both time and material investment. It would be even more interesting to test how parental investment responds to school closure. We have to leave this issue to future research.

References

Alexander, K., Pitcock, S., and Boulay, M. C. (2016). *The summer slide: What we know and can do about summer learning loss*. Teachers College Press.

Alexander, K. L., Entwisle, D. R., and Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational evaluation and policy analysis*, 23(2):171–191.

Alexander, K. L., Entwisle, D. R., and Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American sociological review*, 72(2):167–180.

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, volume 5, pages 111–150.

Ardington, C., Wills, G., and Kotze, J. (2021). Covid-19 learning losses: Early grade reading in south africa. *International Journal of Educational Development*, 86:102480.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). Estimating the production function for human capital: results from a randomized controlled trial in Colombia. *American Economic Review*, 110(1):48–85.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology. General Section*, 28(1):97–104.

Cattan, S., Kamhöfer, D., Karlsson, M., and Nilsson, T. (2017). The short-and long-term effects of student absence: evidence from Sweden. *IZA Discussion Paper*.

Clarke, D. (2021). Rwolf2: Stata module to calculate romano-wolf stepdown p-values for multiple hypothesis testing.

Contini, D., Di Tommaso, M. L., Muratori, C., Piazzalunga, D., and Schiavon, L. (2021). The covid-19 pandemic and school closure: Learning loss in mathematics in primary education. *Working Paper*.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Currie, J. and Almond, D. (2011). Human capital development before age five. In *Handbook of labor economics*, volume 4, pages 1315–1486. Elsevier.

Del Boca, D., Flinn, C., and Wiswall, M. (2014). Household choices and child development. *Review of Economic Studies*, 81(1):137–185.

Dinh, N. T. and Kilenthong, W. T. (2021). Do parental absence and children's gender affect early childhood investment? Evidence from rural Thailand. *The Singapore Economic Review*, 66(5):1443–1468.

Engzell, P., Frey, A., and Verhagen, M. D. (2021). Learning loss due to school closures during the covid-19 pandemic. *Proceedings of the National Academy of Sciences*, 118(17).

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

Gorsuch, R. L. (1983). *Factor Analysis*. Psychology Press, New York.

Halloran, C., Jack, R., Okun, J., and Oster, E. (2021). Pandemic schooling mode and student test scores: Evidence from us states. *Working Paper*.

Hansen, B. (2011). School year length and student performance: Quasi-experimental evidence. *Working Paper*.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.

Ichino, A. and Winter-Ebmer, R. (2004). The long-run educational cost of World War II. *Journal of Labor Economics*, 22(1):57–87.

Isphording, I. E., Lipfert, M., and Pestel, N. (2021). Does re-opening schools contribute to the spread of sars-cov-2? Evidence from staggered summer breaks in Germany. *Journal of Public Economics*, 198:104426.

Jack, W. and Suri, T. (2014). Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution. *American Economic Review*, 104(1):183–223.

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

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, 133(1):97–126.

Knudsen, E. I., Heckman, J. J., Cameron, J. L., and Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the national Academy of Sciences*, 103(27):10155–10162.

Lewis, K., Kuhfeld, M., Ruzek, E., and McEachin, A. (2021). Learning during covid-19: Reading and math achievement in the 2020-21 school year. *Working Paper*.

Maldonado, J. E. and De Witte, K. (2021). The effect of school closures on standardised student test outcomes. *British Educational Research Journal*.

Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.

Romano, J. P. and Wolf, M. (2016). Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. *Statistics & Probability Letters*, 113:38–40.

Sacerdote, B. (2012). When the saints go marching out: Long-term outcomes for student evacuees from hurricanes Katrina and Rita. *American Economic Journal: Applied Economics*, 4(1):109–35.

Schult, J. and Lindner, M. A. (2021). Did students learn less during the covid-19 pandemic? reading and mathematics competencies before and after the first pandemic wave. *Working Paper*.

Solon, G., Haider, S. J., and Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human resources*, 50(2):301–316.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4):518–529.

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. Applied Psychological Measurement, Washington, DC.

Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485):F3–F33.

Todd, P. E. and Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital*, 1(1):91–136.

Tomasik, M. J., Helbling, L. A., and Moser, U. (2020). 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*.

UNICEF et al. (2017). *Overview: MELQO: Measuring Early Learning Quality and Outcomes*. UNESCO Publishing.

A Additional Tables

Table A.1: Share of total variance due to signal and noise for each items

		% signal	% noise
Cognitive skills			
Lieracy			
	Thai letter identification	0.39	0.61
	English letter identification	0.51	0.49
	Word reading	0.34	0.66
Receptive			
	Listening comprehension	0.24	0.76
	Receptive spatial vocabulary	0.20	0.80
Math			
	Mental transformation	0.24	0.76
	Producing a set	0.34	0.66
	Number comparison	0.43	0.57
Working Memory			
	Backward digit span	0.53	0.47
	Forward digit span	0.58	0.42
Non-Cognitive skills			
Non-Cognitive-SDQ			
	Conduct problems	0.59	0.41
	Prosocial behavior	0.39	0.61
	Hyperactive/inattention	0.66	0.34
	Emotional symptoms	0.25	0.75
Non-Cognitive-BPI			
	Headstrong	0.61	0.39
	Anxious/depression	0.35	0.65
	Antisocial	0.47	0.53
	Hyperactive	0.55	0.45
	Dependent	0.45	0.55
	Peer problems	0.31	0.69

Table A.2: Summary statistics for the 2021 Thailand school readiness survey.

	baseline sample			original sample			difference	
	Mean	SD	N	Mean	SD	N	Coeff	p-value
nightlight (prv)	2.459	4.096	7805	2.333	3.988	11478	0.394	0.004
nightlight (dist)	3.012	4.924	7805	2.861	4.831	11478	0.470	0.007
onet (prv)	34.652	1.779	7805	34.527	1.784	11478	0.389	0.000
onet (dist)	35.007	2.452	7805	34.896	2.457	11478	0.346	0.000
std-tch ratio	17.675	7.250	7805	17.512	7.224	11478	0.509	0.073
sch in urban	0.296	0.456	7805	0.288	0.453	11478	0.025	0.162
private sch	0.101	0.301	7805	0.096	0.294	11478	0.015	0.127
kindergarten	0.577	0.494	7805	0.581	0.493	11060	-0.016	0.366
frac of ECE	0.710	0.425	7805	0.711	0.424	10829	-0.003	0.828
frac of GOV	0.673	0.458	7805	0.670	0.458	11086	0.011	0.515
frac of BA tch	0.756	0.414	7805	0.756	0.415	11100	0.001	0.956
frac of MA tch	0.193	0.388	7805	0.195	0.389	11100	-0.008	0.599
Thai lang	0.467	0.499	7805	0.448	0.497	10783	0.069	0.000
Thai nation	0.976	0.152	7805	0.974	0.159	10709	0.008	0.063
care edu: M6	0.177	0.382	7805	0.172	0.377	10370	0.023	0.006
care edu: Voc	0.106	0.307	7805	0.098	0.297	10370	0.032	0.000
care edu: BA+	0.135	0.342	7805	0.122	0.328	10370	0.053	0.000
wealth	0.073	1.129	7805	0.014	1.101	10538	0.229	0.000
no parent	0.179	0.383	7805	0.186	0.390	10569	-0.029	0.003
age	6.369	0.326	7805	6.368	0.327	11478	0.004	0.619
age sq	40.676	4.124	7805	40.662	4.141	11478	0.044	0.639
female	0.481	0.500	7805	0.478	0.500	11478	0.009	0.357
special needs	0.058	0.235	7805	0.060	0.237	10905	-0.005	0.384
sick often	0.081	0.273	7805	0.085	0.279	10698	-0.013	0.041
weight	21.865	5.963	7805	21.768	5.881	11478	0.304	0.011
height	116.51	5.827	7805	116.39	5.808	11478	0.373	0.004
hh size	5.189	1.795	7805	5.155	1.796	9960	0.159	0.001
breakfast days	6.514	1.100	7805	6.535	1.077	10507	-0.082	0.001
read days	1.769	1.859	7805	1.747	1.867	9981	0.100	0.060
care age	42.108	12.216	7805	42.583	12.346	9775	-2.354	0.000
care age sq	1922.3	1120.8	7805	1965.7	1140.6	9775	-215.2	0.000
com-tab own	0.308	0.462	7805	0.289	0.453	10591	0.073	0.000
IT per p head	0.675	0.430	7804	0.659	0.427	9803	0.077	0.000
com-tab p head	0.102	0.202	7804	0.098	0.199	9812	0.023	0.000
tutoring	0.207	0.405	7262	0.209	0.407	9828	-0.009	0.461

Table A.3: Estimation results for robustness checks with an additional mathematics domain.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: Provincial lock-down						
impact	-0.0510	-0.000238	-0.0168	-0.145***	0.0192	0.0232
RW p-value	[0.301]	[0.994]	[0.569]	[0.003]	[0.743]	[0.743]
p-value	(0.160)	(0.991)	(0.481)	(0.000)	(0.702)	(0.490)
No. Obs.	7805	7805	7805	7805	7805	7805
Panel B: School closure						
impact	-0.260**	-0.328***	-0.365***	-0.881***	-0.0719	0.129
RW p-value	[0.023]	[0.006]	[0.005]	[0.001]	[0.658]	[0.393]
p-value	(0.025)	(0.000)	(0.000)	(0.000)	(0.649)	(0.217)
No. Obs.	7489	7489	7489	7489	7489	7489
F-statistics	37.44	37.44	37.44	37.44	37.44	37.44
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.897	0.000	0.004	0.352	0.442	0.669
Panel C: No-school intensity						
impact	-2.641**	-2.908***	-3.347***	-8.779***	-0.528	1.256
RW p-value	[0.015]	[0.002]	[0.001]	[0.001]	[0.755]	[0.462]
p-value	(0.023)	(0.000)	(0.000)	(0.000)	(0.746)	(0.255)
No. Obs.	7489	7489	7489	7489	7489	7489
F-statistics	42.37	42.37	42.37	42.37	42.37	42.37
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.657	0.000	0.000	0.048	0.416	0.545
Panel D: School days						
impact	0.0060**	0.0093***	0.0099***	0.0209***	0.0025	-0.0032
RW p-value	[0.014]	[0.001]	[0.001]	[0.001]	[0.536]	[0.336]
p-value	(0.014)	(0.000)	(0.000)	(0.000)	(0.513)	(0.174)
No. Obs.	7489	7489	7489	7489	7489	7489
F-statistics	46.56	46.56	46.56	46.56	46.56	46.56
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.682	0.004	0.044	0.709	0.509	0.906

Note: RW p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications clustered at classroom level, p-values denote the traditional p-values, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * p<0.10, ** p<0.05, *** p<0.01.

Table A.4: Estimation results for robustness checks when the outcomes are derived using an alternative approach.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: Provincial lock-down						
impact	-0.0744	0.112	0.0254	-0.369***	0.0255	0.0362
RW p-value	[0.513]	[0.151]	[0.650]	[0.001]	[0.751]	[0.703]
p-value	(0.301)	(0.048)	(0.652)	(0.000)	(0.760)	(0.473)
No. Obs.	7805	7805	7805	6114	7503	7804
Panel B: School closure						
impact	-0.173	-0.341	-0.658***	-1.967***	-0.118	0.208
RW p-value	[0.433]	[0.263]	[0.008]	[0.001]	[0.621]	[0.317]
p-value	(0.415)	(0.112)	(0.001)	(0.000)	(0.641)	(0.188)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	37.44	37.44	37.44	27.06	36.86	37.51
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.475	0.000	0.000	0.000	0.467	0.641
Panel C: No-school intensity						
impact	-2.022	-2.316	-5.640**	-19.06***	-0.923	2.019
RW p-value	[0.508]	[0.508]	[0.025]	[0.001]	[0.722]	[0.392]
p-value	(0.365)	(0.272)	(0.005)	(0.000)	(0.723)	(0.225)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	42.37	42.37	42.37	35.83	41.37	42.39
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.533	0.000	0.000	0.000	0.439	0.510
Panel D: School days						
impact	0.0029	0.0125**	0.0194***	0.0401***	0.0041	-0.0052
RW p-value	[0.561]	[0.016]	[0.001]	[0.001]	[0.494]	[0.280]
p-value	(0.537)	(0.009)	(0.000)	(0.000)	(0.493)	(0.146)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	46.56	46.56	46.56	40.91	44.72	46.58
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.389	0.000230	0.00338	0.121	0.552	0.892

Note: RW p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications clustered at classroom level, p-values denote the traditional p-values, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.5: Estimation results for robustness checks with weighting.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: Provincial lock-down						
impact	-0.0279	0.0724	-0.0001	-0.340***	-0.0624	0.0147
RW p-value	[0.886]	[0.449]	[0.995]	[0.004]	[0.657]	[0.772]
p-value	(0.639)	(0.163)	(0.997)	(0.000)	(0.431)	(0.753)
No. Obs.	7805	7805	7805	6114	7503	7804
Panel B: School closure						
impact	-0.0966	-0.0092	-0.253	-1.611*	-0.379	0.0459
RW p-value	[0.884]	[0.962]	[0.248]	[0.059]	[0.230]	[0.737]
p-value	(0.614)	(0.959)	(0.075)	(0.003)	(0.122)	(0.747)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	14.68	14.68	14.68	9.392	14.76	14.70
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.804	0.002	0.000	0.258	0.128	0.558
Panel C: No-school intensity						
impact	-0.995	-0.265	-2.758	-14.31***	-3.969	0.489
RW p-value	[0.842]	[0.897]	[0.145]	[0.001]	[0.214]	[0.729]
p-value	(0.604)	(0.884)	(0.044)	(0.000)	(0.095)	(0.734)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	39.83	39.83	39.83	34.12	40.12	39.84
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.816	0.002	0.000	0.106	0.132	0.570
Panel D: School days						
impact	0.0026	0.0038	0.0101**	0.0367***	0.0115	-0.0016
RW p-value	[0.669]	[0.669]	[0.012]	[0.001]	[0.097]	[0.645]
p-value	(0.582)	(0.414)	(0.003)	(0.000)	(0.044)	(0.646)
No. Obs.	7489	7489	7489	5865	7202	7488
F-statistics	33.70	33.70	33.70	29.29	33.08	33.71
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.932	0.003	0.000	0.820	0.347	0.673

Note: RW p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications clustered at classroom level, p-values denote the traditional p-values, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.6: Estimation results for robustness checks when using the original sample.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: Provincial lock-down						
impact	-0.112*	0.0322	-0.0625*	-0.423***	0.00963	0.0157
RW p-value	[0.085]	[0.404]	[0.085]	[0.001]	[0.904]	[0.904]
p-value	(0.026)	(0.388)	(0.032)	(0.000)	(0.871)	(0.673)
No. Obs.	11478	11478	11478	8894	10773	10823
Panel B: School closure						
impact	-0.238*	-0.111	-0.335***	-1.381***	-0.0727	0.0351
RW p-value	[0.081]	[0.259]	[0.001]	[0.001]	[0.825]	[0.825]
p-value	(0.036)	(0.247)	(0.000)	(0.000)	(0.575)	(0.686)
No. Obs.	10766	10766	10766	8372	10281	10340
F-statistics	58.93	58.93	58.93	46.48	59.78	58.97
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.291	0.001	0.000	0.593	0.287	0.806
Panel C: No-school intensity						
impact	-2.851*	-1.119	-3.769***	-15.61***	-0.792	0.408
RW p-value	[0.072]	[0.309]	[0.001]	[0.001]	[0.855]	[0.855]
p-value	(0.030)	(0.303)	(0.000)	(0.000)	(0.608)	(0.693)
No. Obs.	10766	10766	10766	8372	10281	10340
F-statistics	77.95	77.95	77.95	65.37	76.76	78.65
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.330	0.001	0.000	0.196	0.278	0.790
Panel D: School days						
impact	0.0057*	0.0055*	0.0113***	0.0376***	0.0033	-0.0012
RW p-value	[0.079]	[0.079]	[0.001]	[0.001]	[0.628]	[0.649]
p-value	(0.061)	(0.045)	(0.000)	(0.000)	(0.379)	(0.639)
No. Obs.	10766	10766	10766	8372	10281	10340
F-statistics	72.95	72.95	72.95	62.98	70.29	72.18
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.143	0.004	0.016	0.206	0.377	0.911

Note: RW p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications clustered at classroom level, p-values denote the traditional p-values, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.7: Estimation results for placebo tests.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog teacher	Non-cog parent
Panel A: Provincial lock-down						
impact	0.215**	0.158*	0.0334	0.124	-0.00722	0.0718
RW p-value	[0.028]	[0.070]	[0.529]	[0.237]	[0.915]	[0.402]
p-value	(0.007)	(0.023)	(0.524)	(0.125)	(0.912)	(0.229)
No. Obs.	4360	4360	4360	3357	4316	3252
Panel B: School closure						
impact	0.419**	0.308	0.0650	0.225	-0.0143	0.134
RW p-value	[0.041]	[0.102]	[0.504]	[0.240]	[0.913]	[0.385]
p-value	(0.008)	(0.028)	(0.515)	(0.126)	(0.911)	(0.218)
No. Obs.	4360	4360	4360	3357	4316	3252
F-statistics	27.52	27.52	27.52	30.27	26.42	29.77
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.797	0.799	0.810	0.244	0.643	0.894
Panel C: No-school intensity						
impact	1.106	0.811	0.204	1.060	-0.160	0.297
RW p-value	[0.189]	[0.196]	[0.520]	[0.196]	[0.740]	[0.653]
p-value	(0.042)	(0.063)	(0.507)	(0.060)	(0.707)	(0.392)
No. Obs.	4360	4360	4360	3357	4316	3252
F-statistics	25.78	25.78	25.78	23.01	25.73	26.30
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.272	0.308	0.858	0.718	0.737	0.371
Panel D: School days						
impact	-0.0080	-0.0059	-0.0014	-0.0067	0.0009	-0.0023
RW p-value	[0.114]	[0.114]	[0.490]	[0.114]	[0.752]	[0.551]
p-value	(0.024)	(0.036)	(0.493)	(0.057)	(0.750)	(0.322)
No. Obs.	4360	4360	4360	3357	4316	3252
F-statistics	45.56	45.56	45.56	44.52	45.14	47.66
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.520	0.531	0.957	0.570	0.696	0.485

Note: RW p-values and p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications and the traditional p-values both clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * p<0.10, ** p<0.05, *** p<0.01.

Table A.8: Summary statistics for the 2020 Thailand school readiness survey.

	baseline sample			original sample			difference	
	Mean	SD	N	Mean	SD	N	Coeff	p-value
nightlight (prv)	1.622	1.842	4360	1.746	2.072	8052	-0.271	0.003
nightlight (dist)	2.134	2.782	4360	2.237	2.981	8052	-0.224	0.085
onet (prv)	40.133	4.732	4360	40.231	4.576	8052	-0.213	0.166
onet (dist)	40.332	5.463	4360	40.339	5.410	8052	-0.015	0.937
std-tch ratio	19.766	7.512	4360	19.780	7.525	8037	-0.031	0.908
sch in urban	0.247	0.432	4360	0.253	0.435	8052	-0.011	0.491
private sch	0.146	0.353	4360	0.144	0.352	8052	0.003	0.840
kindergarten	0.908	0.288	4360	0.910	0.286	8052	-0.004	0.720
frac of ECE	0.695	0.431	4360	0.689	0.434	7814	0.012	0.400
frac of GOV	0.629	0.469	4360	0.628	0.470	7787	0.002	0.898
frac of BA tch	0.720	0.436	4360	0.727	0.433	7777	-0.014	0.336
frac of MA tch	0.215	0.405	4360	0.212	0.404	7777	0.007	0.608
Thai language	0.324	0.468	4360	0.327	0.469	7196	-0.008	0.618
Thai nationality	0.983	0.127	4360	0.979	0.143	7185	0.011	0.018
care edu: M6	0.179	0.383	4360	0.160	0.367	7268	0.046	0.000
care edu: Voc	0.103	0.304	4360	0.086	0.281	7268	0.042	0.000
care edu: BA+	0.156	0.363	4360	0.119	0.324	7268	0.093	0.000
wealth	0.051	1.110	4360	-0.020	1.095	6093	0.249	0.000
no parent	3.172	1.167	4360	3.126	1.189	6991	0.123	0.000
age	6.264	0.344	4360	6.276	0.348	8052	-0.026	0.005
age sq	39.358	4.308	4360	39.510	4.357	8052	-0.333	0.005
female	0.493	0.500	4360	0.495	0.500	8052	-0.005	0.660
special needs	0.053	0.225	4360	0.061	0.239	8037	-0.016	0.002
sick often	0.106	0.307	4360	0.120	0.326	7164	-0.038	0.000
weight	20.624	5.169	4360	20.632	5.123	8052	-0.016	0.895
height	114.865	6.059	4360	114.932	6.018	8052	-0.145	0.346
hh size	5.319	1.962	4360	5.314	1.969	6576	0.016	0.782
breakfast days	1.291	2.377	4360	1.335	2.432	6803	-0.125	0.075
read days	1.824	1.945	4360	1.772	1.972	6381	0.164	0.004
care age	41.200	11.871	4360	42.107	12.146	6562	-2.704	0.000
care age sq	1838.3	1082.4	4360	1920.5	1117.7	6562	-245.0	0.000
com-tab own	0.339	0.474	4360	0.301	0.459	6157	0.131	0.000

Table A.9: Heterogeneous effects for school days with respect to household wealth and having tutoring lessons.

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: Interaction between School Days and Household Wealth Index						
school days	0.00244	0.00838**	0.0116***	0.0306***	0.00274	-0.00393
RW p-value	[0.492]	[0.049]	[0.001]	[0.001]	[0.526]	[0.338]
p-value	(0.493)	(0.021)	(0.000)	(0.000)	(0.539)	(0.180)
wealth	-0.457	-0.350	-0.409	-1.003	-0.0787	0.0323
RW p-value	[0.287]	[0.287]	[0.193]	[0.193]	[0.958]	[0.958]
p-value	(0.133)	(0.219)	(0.056)	(0.046)	(0.791)	(0.904)
sch days x wealth	0.00372	0.00253	0.00309	0.00726	0.000748	-0.000193
RW p-value	[0.178]	[0.217]	[0.161]	[0.161]	[0.908]	[0.908]
p-value	(0.081)	(0.200)	(0.039)	(0.039)	(0.718)	(0.917)
F-statistics	31.32	31.32	31.32	30.01	29.84	31.33
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.343	0.00190	0.0194	0.184	0.373	0.939
Panel B: Interaction between School Days and an Indicator for Having Tutoring						
school days	0.00124	0.00710	0.0109***	0.0305***	0.00089	-0.00480
RW p-value	[0.759]	[0.159]	[0.001]	[0.001]	[0.869]	[0.275]
p-value	(0.745)	(0.078)	(0.000)	(0.000)	(0.854)	(0.132)
tutoring	-1.270	-1.109	-1.005	-1.284	-0.282	-0.585
RW p-value	[0.188]	[0.188]	[0.188]	[0.293]	[0.720]	[0.656]
p-value	(0.047)	(0.076)	(0.061)	(0.286)	(0.718)	(0.387)
sch days x tutoring	0.00935	0.00771	0.00706	0.00903	0.00169	0.00357
RW p-value	[0.149]	[0.178]	[0.178]	[0.306]	[0.762]	[0.688]
p-value	(0.035)	(0.074)	(0.056)	(0.274)	(0.754)	(0.444)
F-statistics	25.35	25.35	25.35	20.95	23.85	25.36
Underid. (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Overid. (p-value)	0.651	0.00803	0.0448	0.110	0.768	0.638

Note: RW p-values and p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications and the traditional p-values both clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics, Underid. (p-value) denote p-values for underidentification tests (rank tests), and Overid. (p-value) denote p-values for overidentification tests based on Hansen-J statistics. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * p<0.10, ** p<0.05, *** p<0.01.

B Figures

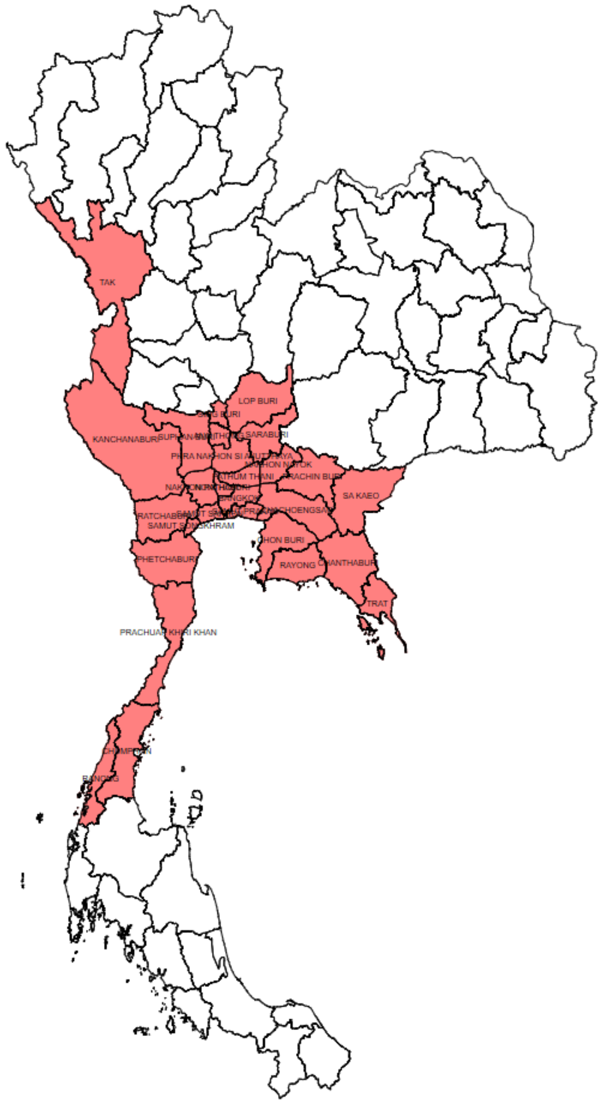


Figure B.1: Map showing 28 lockdowned provinces in red.

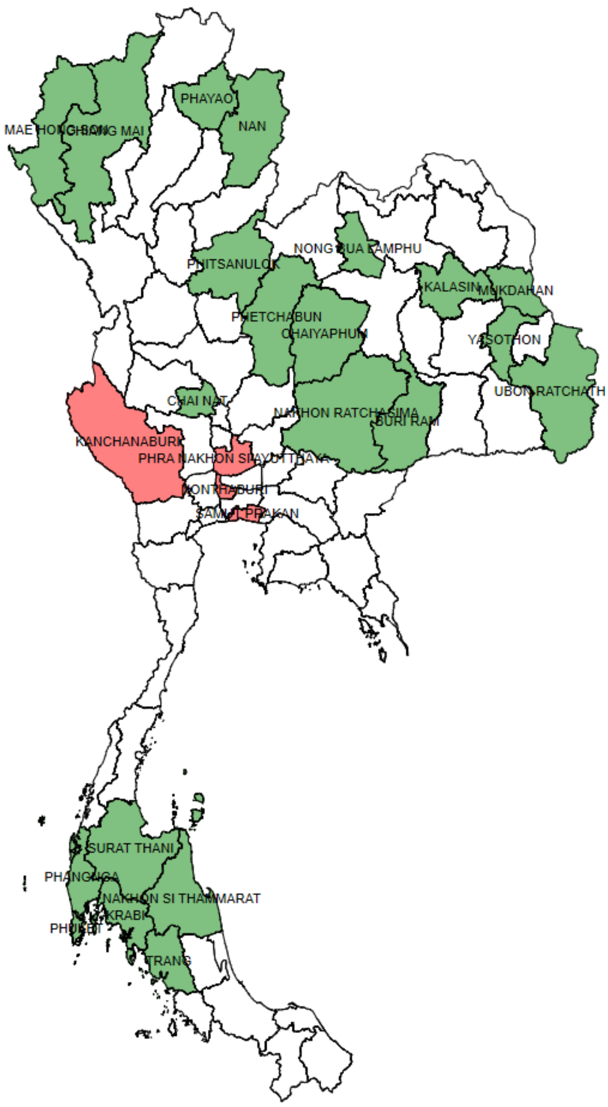


Figure B.2: Map for 25 provinces surveyed in the 2021 Thailand school readiness survey with lockdowned provinces shown in red and no-lockdowned in green.

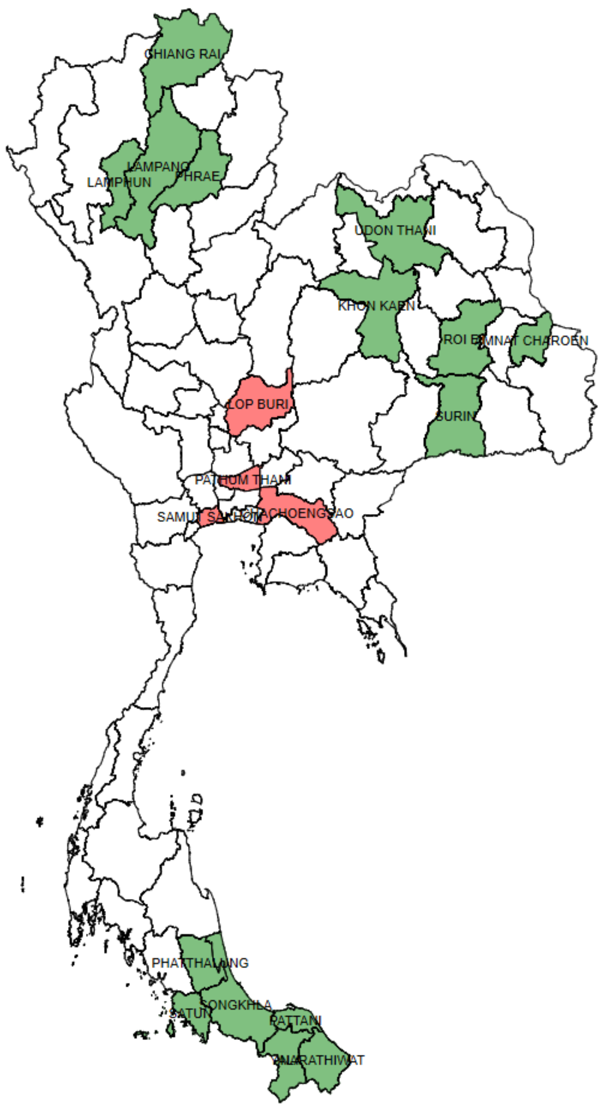


Figure B.3: Map for 19 provinces surveyed in the 2020 Thailand school readiness survey with lockdowned provinces shown in red and no-lockdowned in green.

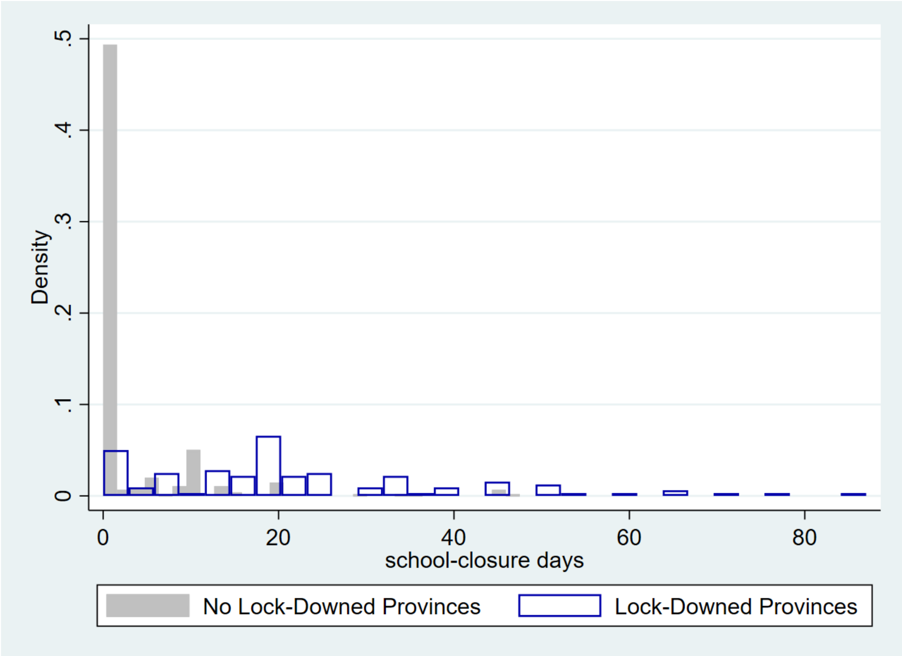


Figure B.4: Histograms of school-closure days from the 2020 Thailand school readiness survey for no-lockdowned and lockdowned provinces.

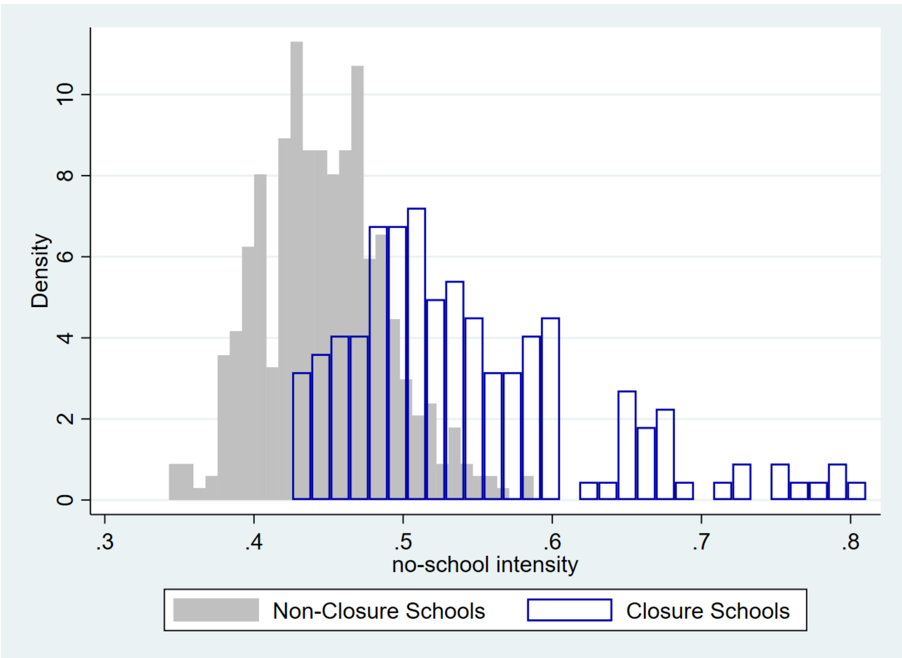


Figure B.5: Histograms of no-school intensity from the 2020 Thailand school readiness survey for non-closure and closure schools.

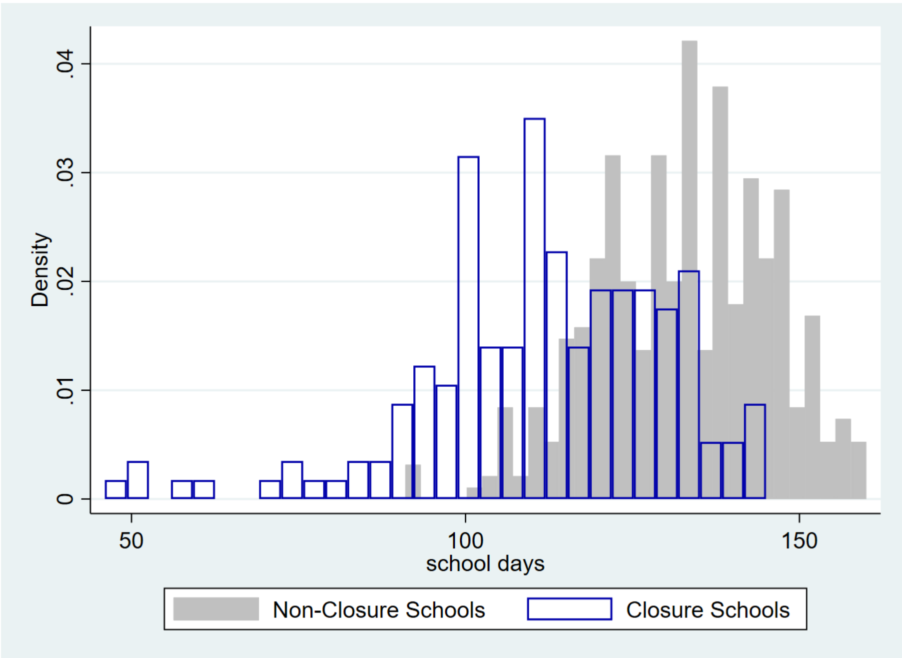


Figure B.6: Histograms of school days from the 2020 Thailand school readiness survey for non-closure and closure schools.

Online Appendix for
Learning Losses from School Closure due to the COVID-19 Pandemic for Thai Kindergartners

Weerachart T. Kilenthong, Khanista Boonsanong,
Sartja Duangchaiyoosook, Wasinee Jantorn and Varunee Khruapradit
Research Instititute for Policy Evaluation and Design (RIPED)
University of the Thai Chamber of Commerce

March 9, 2022

This online appendix contains supplementary materials for the paper titled "Learning Losses from School Closure due to the COVID-19 Pandemic for Thai Kindergartners" by Weerachart T. Kilenthong, Khanista Boonsanong, Sartja Duangchaiyoosook, Wasinee Jantorn and Varunee Kruerpradit.

A Assessment Tools and Questions

This appendix presents key assessment tools and questions, measuring literacy, receptive language, math, working memory, non-cognitive-SDQ and non-cognitive-BPI. Note that assessment tools for literacy, receptive language, math and working memory were adapted from the Measuring Early Learning Quality and Outcomes or MELQO (UNICEF et al., 2017) while non-cognitive-SDQ and non-cognitive-BPI were adapted from the strengths and difficulties questionnaire (SDQ) and the Behavioral Problem Index (BPI), respectively. The original tests were in Thai, which is the only official language of Thailand. All items presented here were translated into English except for Thai letter identification and word reading.

Table A.1: The strengths and difficulties questionnaire (SDQ) used in Thailand School Readiness Survey (TSRS).

Item	Questions	Score		
		Not True	Somewhat True	Certainly True
	Often complains of headaches, ...	2	1	0
Emotional symptoms	Many worries or often seems worried	2	1	0
	Often unhappy, depressed or tearful	2	1	0
	Nervous or clingy in new situations ...	2	1	0
	Many fears, easily scared	2	1	0
	Often loses temper...	2	1	0
Conduct problem	**Generally well behaved ...**	**0**	**1**	**2**
	Often fights with other children ...	2	1	0
	Often lies or cheats	2	1	0
	Steals from home, school ...	2	1	0
Hyperactive	Restless, overactive	2	1	0
	Constantly fidgeting	2	1	0
	Easily distracted...	2	1	0
	Thinks things out before acting	**0**	**1**	**2**
	Good attention span, ...	**0**	**1**	**2**
	Rather solitary, prefers to play alone	2	1	0
Peer problem	**Has at least one good friend**	**0**	**1**	**2**
	Generally liked by other children	**0**	**1**	**2**
	Picked on or bullied....	2	1	0
	Gets along better with adults ...	2	1	0
Prosocial	Considerate of other people's feelings	2	1	0
	Shares readily with other children, ...	2	1	0
	Helpful if someone is hurt ...	2	1	0
	Kind to younger children	2	1	0
	Often volunteers to help others ...	2	1	0

Table A.2: The behavioral problem index questionnaire (BPI) used in Thailand School Readiness Survey (TSRS).

Item	Questions	Score		
		Not True	Sometimes True	Often True
Anxious/ Depressed	Feels/complains no one loves him/her	2	1	0
	Feels worthless or inferior	2	1	0
	Is unhappy, sad, or depressed	2	1	0
Headstrong	Is rather high strung, tense, and nervous	2	1	0
	Is stubborn, sullen, or irritable	2	1	0
	Has strong temper and loses it easily	2	1	0
Antisocial	Bullies or is cruel/mean to others	2	1	0
	Breaks things deliberately	2	1	0
Hyperactive	Is easily confused, seems in a fog	2	1	0
	Is impulsive or acts without thinking	2	1	0
	Is restless, overly active, cannot sit still	2	1	0
Dependent	Cries too much	2	1	0
	Demands a lot of attention	2	1	0
	Is too dependent on others	2	1	0
Peer problems	Is not liked by other children	2	1	0
	Is withdrawn, does not get ... ... involved with others	2	1	0

Direct Assessment

Literacy

Thai letter identification

<u>Ask the child to tell the name of each letter</u>

No.	Thai letter	correct (1)	incorrect (0)	don't know/no answer (-8)
1	ว			
2	ศ			
3	ภ			
4	ฉ			
5	ณ			

English letter identification

<u>Ask the child to tell the name of each letter</u>

No	English letter	correct (1)	incorrect (0)	don't know/no answer (-8)
1	R			
2	K			
3	T			
4	J			
5	V			

Word reading

Ask the child to read each word.

No.	Thai vocabulary	correct (1)	incorrect (0)	don't know/no answer (-8)
1	ไก่			
2	ป่า			
3	เพื่อน			
4	แตงโม			
5	จมูก			

Receptive language

Listening comprehension

<u>Read the following story to the child and ask he/she each question one at a time.</u>

“Once upon a time there was a fat cat. He always wore a red hat. Once when he sleeping, a small mouse came silently and stole the hat. The cat woke up to see his hat gone, got vary angry, and started chasing the mouse. After a while, the mouse was trapped under a table and could not find any way to escape. So the mouse cried to the cat, “Please don’t eat me cat.” If you spare my life, I will return your hat. So, after getting back his hat the cat said, “Never touch my hat again” and he went back to sleep in a happy mood.

No.	Instructions	correct answer	correct (1)	incorrect (0)	don't know/no answer (-8)
1	Who stole the cat’s hat?	The mouse			
2	What was the color of the hat?	Red			
3	Why was the cat chasing the mouse?	Because the mouse took/stole its hat			
4	Where did the cat trap the mouse?	Under the table			
5	Why did the cat decide not to eat the mouse?	Because the mouse gave back the hat			

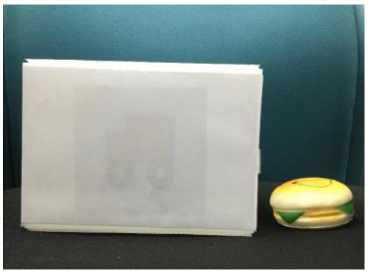
Receptive spatial vocabulary

Ask the child to put the doll to each specified position.

(1)

(2)

(3)

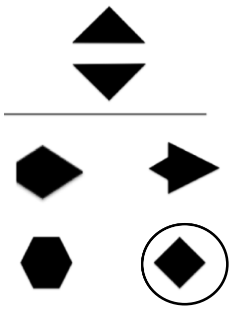
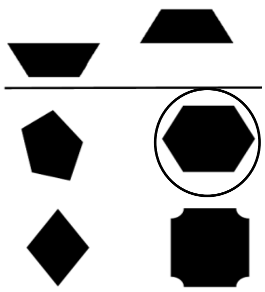
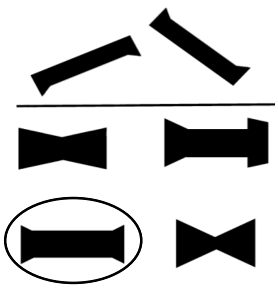


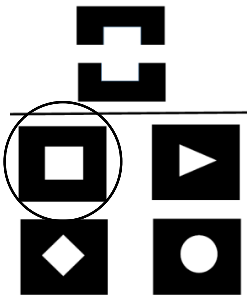
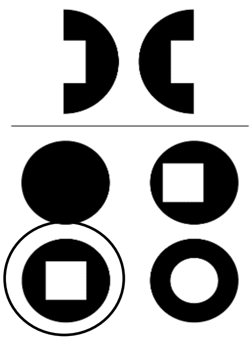
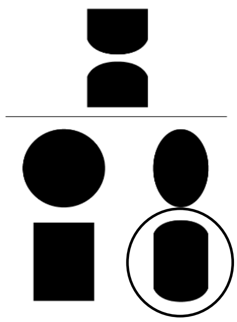
(4)

No.	Instructions	correct (1)	incorrect (0)	don't know/no answer (-8)
1	Put the doll on the box.			
2	Put the doll under the box.			
3	Put the doll in front of the box.			
4	Put the doll next to the box.			

Mathematics

Mental transformation

No.	Instructions	correct answer	correct (1)	incorrect (0)	don't know/no answer (-8)
1	If you put these pieces together (point to set of 2 pieces), they will make one of these shapes (wave hand over 4 choices) Point to the shape the pieces make.				
2	If you put these pieces together (point to set of 2 pieces), they will make one of these shapes (wave hand over 4 choices) Point to the shape the pieces make.				
3	If you put these pieces together (point to set of 2 pieces), they will make one of these shapes (wave hand over 4 choices) Point to the shape the pieces make.				

No.	Instructions	correct answer	correct (1)	incorrect (0)	don't know/no answer (-8)
4	If you put these pieces together (point to set of 2 pieces), they will make one of these shapes (wave hand over 4 choices) Point to the shape the pieces make.				
5	If you put these pieces together (point to set of 2 pieces), they will make one of these shapes (wave hand over 4 choices) Point to the shape the pieces make.				
6	If you put these pieces together (point to set of 2 pieces), they will make one of these shapes (wave hand over 4 choices) Point to the shape the pieces make.				

Producing a Set

No.	Instructions	correct answer	correct (1)	incorrect (0)	don't know/no answer (-8)
1	Please give me three counters.	Hands or pushes over 3 counters			
2	Now, please give me six counters.	Hands or pushes over 6 counters			
3	Now, please give me fourteen counters.	Hands or pushes over 14 counters			

Number Comparison

No.	Instructions	correct answer	correct (1)	incorrect (0)	don't know/no answer (-8)
1	Which number is more/bigger/greater, 3 or 5?	5			
2	Which number is more/bigger/greater, 8 or 6?	8			
3	Which number is smaller/less, 4 or 7?	4			

Symbolic addition

No.	Instructions	child answer	correct (1)	incorrect (0)	don't know/no answer (-8)
1	What is the answer to this? 1 + 2 = ☐				
2	What is the answer to this? 3 + 3 = ☐				
3	What is the answer to this? 6 − 1 = ☐				
4	What is the answer to this? 5 − 2 = ☐				

Working memory

Forward digit span

Show a set of numbers to the child (for 10 seconds) and ask the child to wait for 10 more seconds. Then, ask to child to repeat the set of numbers using the same order. If he/she get it right, then offer a new set of numbers with an additional digit. If he/she get it wrong, repeat the steps with another set of the same number of digits one more time.

No.	Instructions	child answer	correct (1)	incorrect (0)	don't know/no answer (-8)
1	2 6				
2	5 9 0				
3	4 8 6 1				
4	7 3 0 9 4				
5	2 4 9 6 5 8				
6	1 4 6 8 2 4 5				
7	9 0 4 5 6 7 3 1				
8	1 4 3 6 7 8 9 0 2				
9	9 1 5 4 3 8 7 6 0 2				

Backward digit span

<u>Show a set of numbers to the child (for 10 seconds) and ask the child to wait for 10 more seconds. Then, ask to child to repeat the set of numbers using the reversed order. If he/she get it right, then offer a new set of numbers with an additional digit. If he/she get it wrong, repeat the steps with another set of the same number of digits one more time.</u>

No.	Instructions	child answer	correct (1)	incorrect (0)	don't know/no answer (-8)
1	4 8				
2	5 8 2				
3	6 8 9 3				
4	5 1 3 7 4				
5	1 7 3 6 2 8				
6	7 9 0 4 8 6 1				
7	2 8 3 9 4 0 6 5				
8	9 4 0 5 8 2 6 7 1				
9	5 6 3 0 1 8 2 9 4 7				

B Standardized Scores using Kernel-weighted Local Polynomial Smoothing

This appendix summarizes briefly how age-standardized scores were generated using kernel-weighted local polynomial smoothing (KLPS). Following Attanasio et al. (2020), this method can be divided into 3 steps.

The first step is to estimate the following polynomial equations, for each skill s and item j separately,

$$y_{ij}^{s} = f_{j}^{s}\left(X_{i}\right) + \varepsilon_{ij}^{s} \tag{A.1}$$

using kernel-weighted local polynomial smoothing methods, where X_i is the age of the individual, function $f_j^s\left(X_i\right)$ is a third-degree polynomial function, and y_{ij}^s is a raw score for item j within a skill domain s of child i. The age-conditional mean, $\hat{y}_{ij}^s$, is the predicted value of the outcome from (A.1).

The second step is to estimate the age-conditional standard deviation by regressing the square of the estimated residuals from (A.1) on a third-degree polynomial function:

$$\left(y_{ij}^{s} - \hat{y}_{ij}^{s}\right)^{2} = g_{j}^{s}\left(X_{i}\right) + \epsilon_{ij}^{s}, \tag{A.2}$$

again using kernel-weighted local polynomial smoothing methods, where X_i is the age of the individual and function $g_j^s\left(X_i\right)$ is a third-degree polynomial function. We can then estimate age-conditional standard deviation, $\hat{\sigma}_{ij}^s$, as the square root of the predicted value of the outcome from (A.2).

The third step is to compute age-standardized scores as follows:

$$Y_{ij}^{s} = \frac{y_{ij}^{s} - \hat{y}_{ij}^{s}}{\hat{\sigma}_{ij}^{s}}. \tag{A.3}$$

This procedure should result in smoothly distributed internally standardized scores, with mean zero and standard deviation one conditional on child age.

C Empirical Results for Exploratory Factor Analysis (EFA)

This appendix presents empirical results from an exploratory factor analysis (EFA) using the age-standardized scores as raw scores. We perform an EFA for each of the following four groups of items separately

1. language consists of 5 items including Thai letter identification, English letter identification, word reading, receptive spatial vocabulary, listening comprehension;

2. mathematics consists of 3 items including number comparison, producing a set, mental transformation;

3. non-cognitive-SDQ consists of 5 items including conduct problems, emotional symptoms, hyperactivity/inattention, prosocial behaviour, peer problem;

4. non-cognitive-BPI consists of 6 items including headstrong, anxiousness/depression, antisocial, hyperactive, dependent, peer problems.

We determine the number of latent factors for each group using the eigenvalue criteria (eigenvalue is larger than one) and factor loading after performing factor rotation with quartimin rotation. In particular, we will disregard an item whose factor loading after the quartimin rotation is less than 0.3, following Attanasio et al. (2020). For the first group, we decided to retain two latent factors, namely literacy and receptive language even though eigenvalue of the second one is less than one. We keep this factor because its two items (listening comprehension, receptive spatial vocabulary) are important skills for young children. See table A.3. For mathematics, it is clear that there is only one latent factor. See table A.4. For non-cognitive-SDQ, there seem to have two factors whose eigenvalues are both larger than one. However, the factor loading for the emotional symptoms item is larger for the first factor, and, therefore, we assign it to the first factor. That leaves the second factor with only one item, the peer problem, which is then dropped. See table A.5. Similarly, for non-cognitive-BPI, there are two factors whose eigenvalues are larger than one. However, the second factor has only one item whose loading is larger than 0.3. Therefore, we assign all six items to the first factor. See table A.6. Note that we cannot perform this analysis for working memory since it has only two items.

In conclusion, there are five latent factors, namely literacy, receptive language, math, working memory, non-cognitive-SDQ and non-cognitive-BPI.

D Empirical Results for Confirmatory Factor Analysis (CFA)

The dedicated measurement system for all five latent factors, namely literacy, receptive language, math, working memory, non-cognitive-SDQ and non-cognitive-BPI, are as

Table A.3: Factor loadings and eigenvalues for language items including Thai letter identification, English letter identification, word reading, receptive spatial vocabulary, listening comprehension.

Item	factor loadings for	
	factor 1	factor 2
Thai letter identification	0.5949	
English letter identification	0.7447	
word reading	0.5762	
receptive spatial vocabulary		0.4647
listening comprehension		0.3550
eigenvalue	1.3064	0.4069

Table A.4: Factor loadings and eigenvalues for mathematics items including number comparison, producing a set, mental tranformation.

Item	factor loadings for
	factor 1
number comparison	0.4978
producing a set	0.6045
mental tranformation	0.6277
eigenvalue	1.0072

Table A.5: Factor loadings and eigenvalues for non-cognitive-SDQ items including conduct problems, emotional symptoms, hyperactivity/inattention, prosocial behaviour, peer problem.

Item	factor loadings for	
	factor 1	factor 2
conduct problems	0.7297	
emotional symptoms	0.5674	
hyperactivity/inattention	0.8068	
prosocial behaviour	0.4238	0.3476
peer problem		0.9679
eigenvalue	1.7480	1.2224

Table A.6: Factor loadings and eigenvalues for non-cognitive-BPI items including headstrong, anxiousness/depression, antisocial, hyperactive, dependent, peer problems.

Item	factor loadings for	
	factor 1	factor 2
headstrong	0.7300	
anxiousness/depression	0.3668	0.9303
antisocial	0.6852	
hyperactive	0.7343	
dependent	0.6581	
peer problems	0.5093	
eigenvalue	2.3685	1.0418

follows.

$$Y_j^s = \alpha_j^s + \lambda_j^s \theta^s + \varepsilon_j^s, \text{ for } j = 1 \ldots, J_s \tag{A.4}$$

where Y_j^s is an age-standardized score of item j for latent factor s, θ^s is a latent factor s, λ_j^s is a factor loading of item j for factor s, and ε_j^s is a mean zero measurement error term which is assumed to be independent of the latent factors and each other. There are five latent factors or skills in our case, namely literacy ($J_s = 3$ including Thai letter identification, English letter identification, word reading), receptive language ($J_s = 2$ including listening comprehension, receptive spatial vocabulary), math ($J_s = 3$ including mental transformation, number comparison, producing a set), working memory ($J_s = 2$ including backward and forward digit spans), non-cognitive-SDQ ($J_s = 4$ including conduct problems, emotional symptoms, hyperactivity/inattention, prosocial behaviour) and non-cognitive-BPI ($J_s = 6$ including headstrong, anxiousness/depression, antisocial, hyperactive, dependent, peer problems). All of these latent factors are freely correlated with each other.

The factor model is estimated using a confirmatory factor analysis (CFA) approach (e.g., Gorsuch, 1983; Thompson, 2004) using the full-information maximum likelihood estimation with normally distributed errors. Under this approach, missing values are assumed to be missing at random and all observed and latent variables are assumed to be distributed jointly normal. Following Anderson and Rubin (1956), we normalize the factor model by setting its scale in such a way that the factor loading on the first item of each factor is one; that is, $\lambda_1^s = 1$ for all s. In particular, the normalization measures for each five factors are Thai letter identification, listening comprehension, mental transformation, backward digit span, conduct problems, and headstrong, respectively. See Cunha et al. (2010) for a general identification.

Table A.7 presents factor loadings and fit statistics for the factor model (A.4). The relevant fit statistics, including Tucker-Lewis index (TLI), Comparative fit index (CFI) and Root mean squared error of approximation (RMSEA), indicate that the model has a good fit.

Table A.8 presents variance-covariance matrix of latent factors. All cognitive factors, namely literacy, receptive language, math and working memory, are highly correlated while the correlation between the non-cognitive factors, namely non-cognitive-SDQ and non-cognitive-BPI, are much weaker. In addition, the correlations between cognitive and non-cognitive items are also relatively weak.

Technically, we implement the CFA using an sem command in STATA16, whose the only factor score prediction option is a regression-based method. We, therefore, have to calculate Bartlett factor scores manually. For comparison, we present kernel densities of factor scores from both the STATA-default regression-based method and manually-calculated Bartlett in figure A.1-A.6. Note that Bartlett and regression-based factor scores are highly correlated for all latent factors except receptive language.

Table A.7: Estimates of Factor Loadings for Measurement System

	Coefficient	Standard Error
Literacy		
Thai letter identification	1	
English letter identification	1.146***	(0.024)
Word reading	0.938***	(0.021)
Receptive Language		
Listening comprehension	1	
Receptive spatial vocabulary	0.916***	(0.032)
Math		
Mental transformation	1	
Producing a set	1.178***	(0.030)
Number comparison	1.331***	(0.032)
Working Memory		
Backward digit span	1	
Forward digit span	1.055***	(0.028)
Non-Cognitive-SDQ		
Conduct problems	1	
Prosocial behavior	0.809***	(0.013)
Hyperactive/inattention	1.053***	(0.015)
Emotional symptoms	0.655***	(0.013)
Non-Cognitive-BPI		
Headstrong	1	
Anxious/depression	0.760***	(0.012)
Antisocial	0.879***	(0.012)
Hyperactive	0.949***	(0.013)
Dependent	0.853***	(0.012)
Peer problems	0.714***	(0.013)
Sample size	12,331	
Tucker-Lewis index (TLI)	0.965	
Comparative fit index (CFI)	0.972	
Root mean squared error of approximation (RMSEA)	0.029	

Note: Standard errors are in parenthesis and * p<0.10, ** p<0.05, *** p<0.01. TLI (Tucker and Lewis, 1973) and CFI (Bentler, 1990a,b) closes to one indicating good fit while RMSEA (Browne and Cudeck, 1992; Steiger, 1990) closes to zero showing good fit.

Table A.8: Estimates of Correlation among Latent Factors

	Literacy	Receptive Language	Math	Working Memory	Non-cog SDQ	Non-cog BPI
Literacy	1.000					
Receptive language	0.601***	1.000				
	(0.000)					
Math	0.725***	0.962***	1.000			
	(0.000)	(0.000)				
Working memory	0.612***	0.770***	0.812***	1.000		
	(0.000)	(0.000)	(0.000)			
Non-cognitive-SDQ	0.296***	0.299***	0.392***	0.262***	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)		
Non-cognitive-BPI	0.134***	0.052***	0.131***	0.099***	0.258***	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Note: P-values are in parenthesis and * p<0.10, ** p<0.05, *** p<0.01.

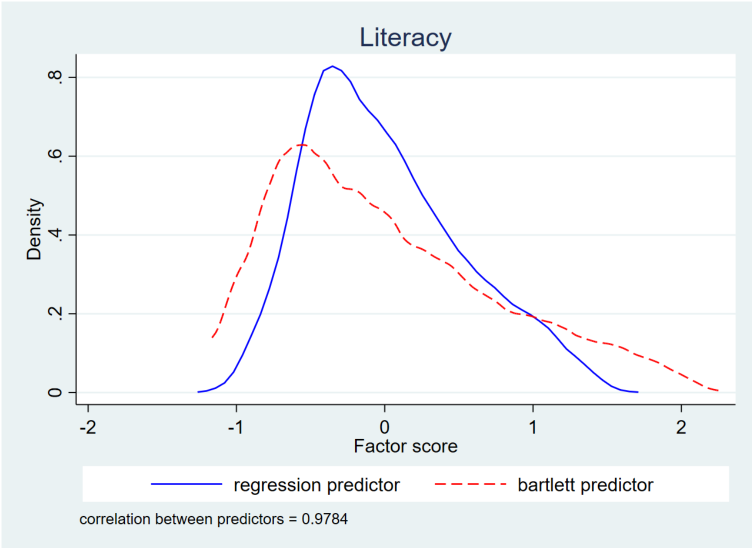


Figure A.1: The distribution of literacy using Bartlett method (red dash line) and regression-based method (blue bold line).

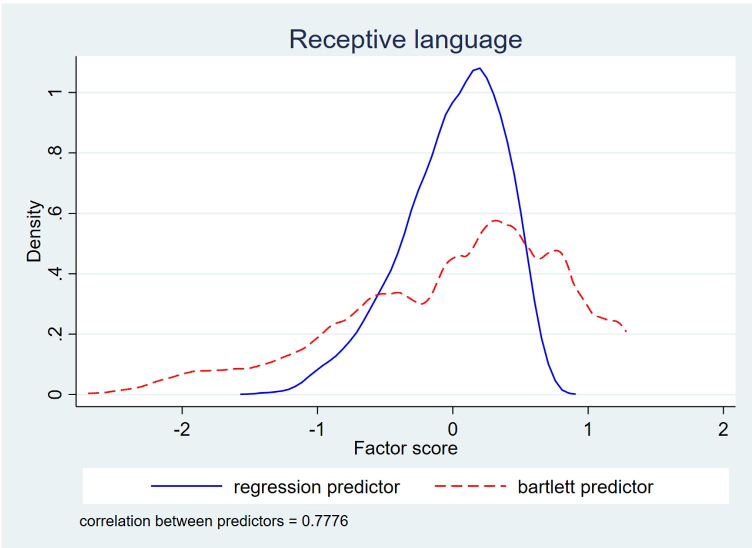


Figure A.2: The distribution of receptive language using Bartlett method (red dash line) and regression-based method (blue bold line).

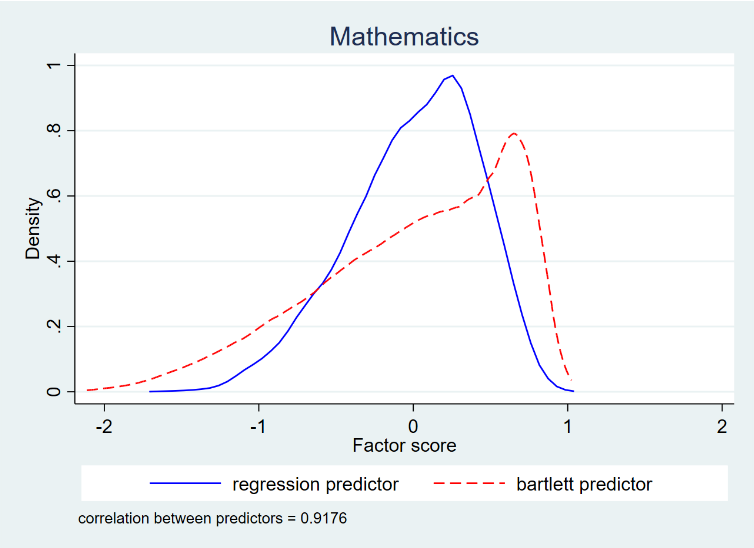


Figure A.3: The distribution of mathematics using Bartlett method (red dash line) and regression-based method (blue bold line).

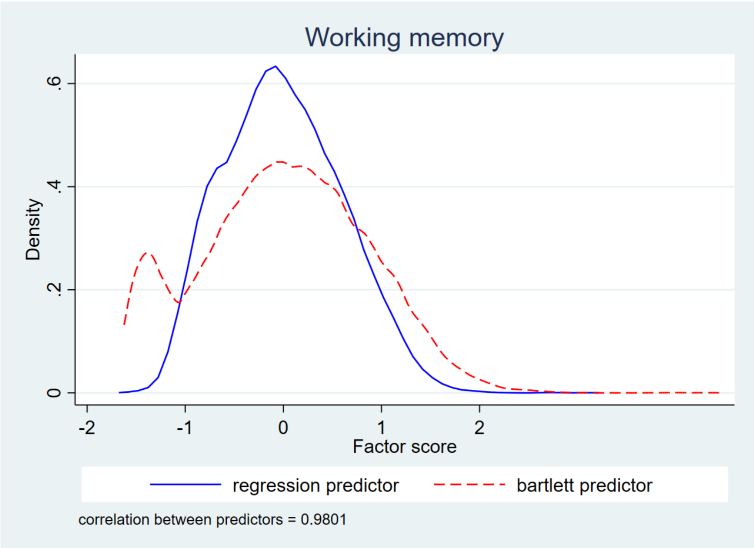


Figure A.4: The distribution of working memory using Bartlett method (red dash line) and regression-based method (blue bold line).

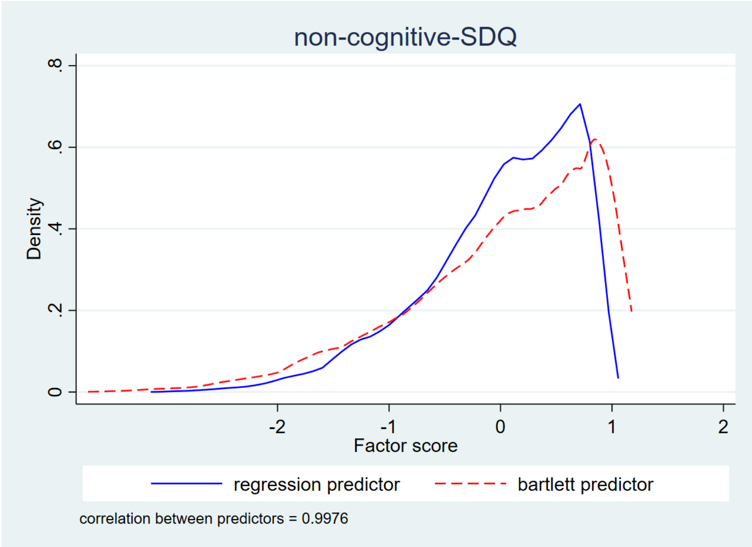


Figure A.5: The distribution of non-cognitive-SDQ using Bartlett method (red dash line) and regression-based method (blue bold line).

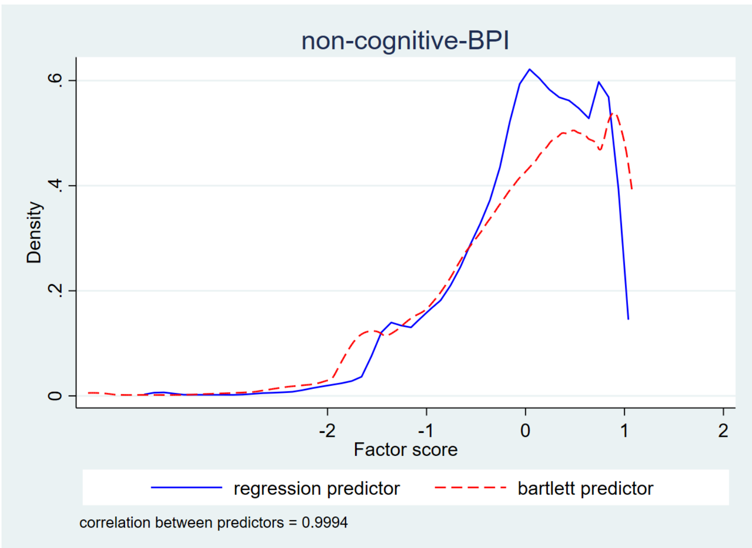


Figure A.6: The distribution of non-cognitive-BPI using Bartlett method (red dash line) and regression-based method (blue bold line).

E First-Stage Regression Results for Instrumental Variable Estimations

This section presents first-stage regression results corresponding to the IV estimates in table 1-3 in section 4 of the main text.

Table A.9: First-stage estimation results for school closure

	(1)	(2)	(3)	(4)
provincial lock-down	0.217*** (0.000)	0.190*** (0.000)	0.217*** (0.000)	0.217*** (0.000)
Covid cases (prv)	-0.00123*** (0.000)	-0.00117*** (0.000)	-0.00121*** (0.000)	-0.00124*** (0.000)
std-tch ratio	0.00113 (0.598)	0.00063 (0.774)	0.00054 (0.803)	0.00112 (0.601)
sch in urban	-0.00716 (0.853)	-0.00806 (0.839)	-0.0167 (0.668)	-0.00701 (0.856)
private sch	0.138* (0.033)	0.143* (0.026)	0.140* (0.033)	0.138* (0.033)
kindergarten	-0.0112 (0.716)	-0.0228 (0.475)	-0.0122 (0.692)	-0.0110 (0.721)
frac of ECE	0.0320 (0.410)	0.0343 (0.397)	0.0365 (0.352)	0.0318 (0.412)
frac of GOV	-0.0212 (0.592)	-0.0122 (0.760)	-0.0225 (0.574)	-0.0213 (0.590)
frac of BA tch	0.0570 (0.455)	0.0472 (0.545)	0.0581 (0.455)	0.0570 (0.455)
frac of MA tch	0.0967 (0.251)	0.0846 (0.326)	0.0918 (0.284)	0.0969 (0.250)
Thai language	0.257*** (0.000)	0.254*** (0.000)	0.260*** (0.000)	0.257*** (0.000)
Thai nationality	0.0138 (0.806)	0.0427 (0.470)	0.0225 (0.678)	0.0137 (0.807)
care edu: M6	-0.0103 (0.531)	-0.0224 (0.220)	-0.00909 (0.581)	-0.0103 (0.530)
care edu: Voc	0.0149 (0.420)	-0.000961 (0.962)	0.0163 (0.384)	0.0149 (0.423)
care edu: BA+	-0.0121 (0.546)	-0.00904 (0.683)	-0.0160 (0.433)	-0.0123 (0.541)
wealth	0.00501 (0.574)	0.00642 (0.500)	0.00662 (0.460)	0.00511 (0.566)
no parent	-0.0256 (0.086)	-0.0361* (0.024)	-0.0205 (0.182)	-0.0261 (0.081)
age	-0.702 (0.224)	-0.476 (0.451)	-0.688 (0.246)	-0.709 (0.221)
age sq	0.0548 (0.232)	0.0366 (0.465)	0.0545 (0.246)	0.0553 (0.228)
female	0.0185 (0.065)	0.0242* (0.030)	0.0165 (0.113)	0.0186 (0.063)
special needs	0.0137 (0.596)	0.00769 (0.815)	0.0156 (0.556)	0.0121 (0.640)
sick often	0.00391 (0.831)	0.0137 (0.524)	0.00387 (0.837)	0.00279 (0.879)
weight	-0.000737 (0.582)	-0.000996 (0.488)	-0.000802 (0.556)	-0.000737 (0.582)
height	0.00113 (0.485)	0.00256 (0.131)	0.000979 (0.552)	0.00114 (0.482)
hh size	-0.00552 (0.087)	-0.00335 (0.351)	-0.00524 (0.109)	-0.00557 (0.085)
breakfast days	-0.00958* (0.048)	-0.0135* (0.012)	-0.00874 (0.078)	-0.00963* (0.047)
read days	-0.00245 (0.434)	-0.00190 (0.570)	-0.00198 (0.532)	-0.00246 (0.431)
care age	-0.00784** (0.006)	-0.0105*** (0.001)	-0.00788** (0.007)	-0.00783** (0.006)
care age sq	0.00009** (0.004)	0.00012*** (0.000)	0.00009** (0.005)	0.00009** (0.004)
com-tab own	-0.00746 (0.613)	-0.0128 (0.426)	-0.00568 (0.705)	-0.00749 (0.612)
constant	2.543 (0.169)	1.807 (0.369)	2.495 (0.189)	2.568 (0.166)
F-stat	37.440	27.060	36.864	37.510
No. Obs.	7489	5865	7202	7488

Note: Stars are based on p-values with * p<0.10, ** p<0.05, *** p<0.01 and clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics.

Table A.10: First-stage estimation results for no-school intensity

	(1)	(2)	(3)	(4)
provincial lock-down	0.0236***	0.0235***	0.0235***	0.0236***
	(0.000)	(0.000)	(0.000)	(0.000)
Covid cases (prv)	-0.00010***	-0.00010***	-0.00010***	-0.00010***
	(0.000)	(0.000)	(0.000)	(0.000)
std-tch ratio	-0.000246	-0.000261	-0.000255	-0.000247
	(0.064)	(0.052)	(0.059)	(0.063)
sch in urban	-0.000414	-0.000540	-0.000632	-0.000404
	(0.868)	(0.832)	(0.804)	(0.872)
private sch	-0.00307	-0.00145	-0.00293	-0.00308
	(0.433)	(0.702)	(0.460)	(0.432)
kindergarten	-0.00036	-0.00005	-0.00033	-0.00035
	(0.860)	(0.981)	(0.874)	(0.865)
frac of ECE	0.00358	0.00364	0.00373	0.00357
	(0.124)	(0.127)	(0.114)	(0.125)
frac of GOV	0.00010	0.00174	0.00003	0.00009
	(0.969)	(0.510)	(0.991)	(0.972)
frac of BA tch	-0.00526	-0.00548	-0.00570	-0.00526
	(0.301)	(0.285)	(0.255)	(0.301)
frac of MA tch	-0.00461	-0.00574	-0.00515	-0.00459
	(0.412)	(0.310)	(0.354)	(0.413)
Thai language	0.0136***	0.0127***	0.0135***	0.0136***
	(0.000)	(0.000)	(0.000)	(0.000)
Thai nationality	0.00307	0.00366	0.00333	0.00307
	(0.428)	(0.439)	(0.392)	(0.429)
care edu: M6	-0.000961	-0.00144	-0.000726	-0.000963
	(0.366)	(0.222)	(0.500)	(0.365)
care edu: Voc	0.00096	-0.00009	0.00143	0.00096
	(0.417)	(0.943)	(0.237)	(0.420)
care edu: BA+	-0.00248	-0.00226	-0.00280*	-0.00249
	(0.062)	(0.111)	(0.037)	(0.061)
wealth	0.00034	0.00067	0.00031	0.00035
	(0.536)	(0.248)	(0.575)	(0.528)
no parent	-0.00063	-0.00135	-0.00019	-0.00066
	(0.583)	(0.242)	(0.870)	(0.562)
age	0.0183	0.0319	0.0220	0.0178
	(0.603)	(0.400)	(0.542)	(0.613)
age sq	-0.00170	-0.00285	-0.00196	-0.00166
	(0.545)	(0.344)	(0.495)	(0.555)
female	0.00082	0.00100	0.00092	0.00082
	(0.220)	(0.165)	(0.175)	(0.216)
special needs	-0.00303*	-0.00314	-0.00305*	-0.00314*
	(0.050)	(0.105)	(0.049)	(0.042)
sick often	0.00153	0.00185	0.00127	0.00145
	(0.204)	(0.175)	(0.302)	(0.229)
weight	0.00016	0.00015	0.00017	0.00016
	(0.091)	(0.142)	(0.081)	(0.091)
height	-0.00011	-0.00003	-0.00012	-0.00011
	(0.311)	(0.802)	(0.274)	(0.314)
hh size	-0.00023	-0.00016	-0.00020	-0.00023
	(0.310)	(0.514)	(0.377)	(0.304)
breakfast days	-0.00094**	-0.00113**	-0.00086**	-0.00094**
	(0.003)	(0.002)	(0.009)	(0.003)
read days	-0.00004	0.00018	0.00001	-0.00004
	(0.851)	(0.372)	(0.941)	(0.846)
care age	0.00001	-0.00017	-0.00003	0.00001
	(0.955)	(0.399)	(0.888)	(0.955)
care age sq	0.000000298	0.00000267	0.000000624	0.000000292
	(0.882)	(0.229)	(0.764)	(0.885)
com-tab own	-0.00188*	-0.00300**	-0.00172	-0.00188*
	(0.048)	(0.005)	(0.066)	(0.048)
constant	0.555***	0.505***	0.548***	0.557***
	(0.000)	(0.000)	(0.000)	(0.000)
F-stat	42.365	35.827	41.370	42.394
No. Obs.	7489	5865	7202	7488

Note: Stars are based on p-values with * p<0.10, ** p<0.05, *** p<0.01 and clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics.

Table A.11: First-stage estimation results for log of no-school intensity

	(1)	(2)	(3)	(4)
provincial lock-down	0.0404***	0.0401***	0.0403***	0.0404***
	(0.000)	(0.000)	(0.000)	(0.000)
Covid cases (prv)	-0.00018***	-0.00017***	-0.00018***	-0.00018***
	(0.000)	(0.000)	(0.000)	(0.000)
std-tch ratio	-0.00043	-0.00049	-0.00045	-0.00043
	(0.063)	(0.052)	(0.059)	(0.062)
sch in urban	-0.00094	-0.00111	-0.00135	-0.00092
	(0.829)	(0.802)	(0.760)	(0.832)
private sch	-0.00464	-0.00185	-0.00434	-0.00465
	(0.503)	(0.784)	(0.535)	(0.502)
kindergarten	-0.00051	0.00007	-0.00046	-0.00048
	(0.887)	(0.985)	(0.898)	(0.893)
frac of ECE	0.00626	0.00632	0.00655	0.00625
	(0.126)	(0.134)	(0.115)	(0.127)
frac of GOV	0.00033	0.00315	0.00024	0.00032
	(0.943)	(0.499)	(0.960)	(0.945)
frac of BA tch	-0.00950	-0.00973	-0.0103	-0.00950
	(0.282)	(0.277)	(0.237)	(0.282)
frac of MA tch	-0.00840	-0.0101	-0.00938	-0.00837
	(0.389)	(0.304)	(0.332)	(0.390)
Thai language	0.0238***	0.0223***	0.0235***	0.0238***
	(0.000)	(0.000)	(0.000)	(0.000)
Thai nationality	0.00584	0.00717	0.00631	0.00583
	(0.419)	(0.421)	(0.385)	(0.420)
care edu: M6	-0.00155	-0.00246	-0.00114	-0.00155
	(0.406)	(0.236)	(0.545)	(0.405)
care edu: Voc	0.00178	-0.00004	0.00261	0.00177
	(0.391)	(0.986)	(0.218)	(0.394)
care edu: BA+	-0.00424	-0.00396	-0.00481*	-0.00426
	(0.070)	(0.114)	(0.042)	(0.069)
wealth	0.000552	0.00113	0.000514	0.000564
	(0.562)	(0.259)	(0.596)	(0.553)
no parent	-0.00112	-0.00238	-0.00038	-0.00118
	(0.566)	(0.232)	(0.852)	(0.545)
age	0.0304	0.0514	0.0369	0.0296
	(0.619)	(0.434)	(0.556)	(0.629)
age sq	-0.00287	-0.00468	-0.00334	-0.00281
	(0.556)	(0.372)	(0.504)	(0.565)
female	0.00145	0.00176	0.00162	0.00146
	(0.205)	(0.161)	(0.165)	(0.201)
special needs	-0.00505	-0.00508	-0.00508	-0.00524
	(0.063)	(0.135)	(0.063)	(0.054)
sick often	0.00267	0.00325	0.00219	0.00254
	(0.198)	(0.167)	(0.303)	(0.223)
weight	0.000269	0.000232	0.000281	0.000269
	(0.102)	(0.172)	(0.093)	(0.102)
height	-0.00017	-0.00002	-0.00019	-0.00017
	(0.363)	(0.917)	(0.321)	(0.366)
hh size	-0.00039	-0.00028	-0.00034	-0.00039
	(0.317)	(0.516)	(0.385)	(0.311)
breakfast days	-0.00167**	-0.00197**	-0.00153**	-0.00168**
	(0.002)	(0.002)	(0.007)	(0.002)
read days	-0.0000509	0.000333	0.0000366	-0.0000530
	(0.881)	(0.351)	(0.916)	(0.876)
care age	-0.00002	-0.00032	-0.00009	-0.00002
	(0.961)	(0.357)	(0.792)	(0.961)
care age sq	0.000000928	0.00000495	0.00000156	0.000000917
	(0.789)	(0.197)	(0.663)	(0.791)
com-tab own	-0.00327*	-0.00519**	-0.00302	-0.00327*
	(0.049)	(0.005)	(0.067)	(0.049)
constant	-0.591**	-0.672**	-0.603**	-0.588**
	(0.003)	(0.001)	(0.003)	(0.003)
F-stat	45.793	38.890	44.713	45.825
No. Obs.	7489	5865	7202	7488

Note: Stars are based on p-values with * p<0.10, ** p<0.05, *** p<0.01 and clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics.

Table A.12: First-stage estimation results for school days

	(1)	(2)	(3)	(4)
provincial lock-down	-6.734***	-6.888***	-6.565***	-6.734***
	(0.000)	(0.000)	(0.000)	(0.000)
Covid cases (prv)	0.0643***	0.0630***	0.0650***	0.0643***
	(0.000)	(0.000)	(0.000)	(0.000)
std-tch ratio	0.100	0.118	0.0993	0.100
	(0.122)	(0.075)	(0.133)	(0.122)
sch in urban	1.285	1.106	1.463	1.283
	(0.277)	(0.361)	(0.224)	(0.278)
private sch	-0.351	-1.125	-0.539	-0.350
	(0.841)	(0.504)	(0.762)	(0.842)
kindergarten	0.409	0.325	0.430	0.406
	(0.679)	(0.752)	(0.665)	(0.681)
frac of ECE	-1.579	-1.577	-1.669	-1.577
	(0.173)	(0.195)	(0.156)	(0.174)
frac of GOV	-0.0463	-0.350	0.00515	-0.0445
	(0.970)	(0.781)	(0.997)	(0.971)
frac of BA tch	3.558	3.330	3.629	3.558
	(0.139)	(0.174)	(0.129)	(0.139)
frac of MA tch	2.965	2.982	3.037	2.961
	(0.268)	(0.274)	(0.256)	(0.268)
Thai language	-4.733***	-4.586***	-4.650***	-4.735***
	(0.000)	(0.000)	(0.000)	(0.000)
Thai nationality	0.662	0.803	0.489	0.663
	(0.698)	(0.678)	(0.775)	(0.698)
care edu: M6	0.268	0.369	0.163	0.268
	(0.603)	(0.530)	(0.753)	(0.602)
care edu: Voc	-0.787	-0.569	-0.967	-0.786
	(0.164)	(0.361)	(0.094)	(0.164)
care edu: BA+	0.811	0.831	0.923	0.814
	(0.210)	(0.238)	(0.157)	(0.209)
wealth	-0.117	-0.291	-0.0612	-0.118
	(0.640)	(0.271)	(0.811)	(0.635)
no parent	0.0324	0.527	-0.180	0.0404
	(0.953)	(0.349)	(0.747)	(0.941)
age	-25.89	-26.19	-27.77	-25.79
	(0.135)	(0.155)	(0.118)	(0.137)
age sq	2.280	2.335	2.421	2.272
	(0.098)	(0.111)	(0.086)	(0.099)
female	-0.0758	-0.0948	-0.150	-0.0773
	(0.805)	(0.780)	(0.632)	(0.801)
special needs	0.819	1.321	0.835	0.844
	(0.274)	(0.180)	(0.263)	(0.261)
sick often	-0.494	-0.591	-0.404	-0.476
	(0.385)	(0.362)	(0.482)	(0.403)
weight	-0.0928*	-0.0881	-0.0977*	-0.0928*
	(0.041)	(0.059)	(0.035)	(0.041)
height	0.0742	0.0469	0.0724	0.0741
	(0.151)	(0.390)	(0.168)	(0.151)
hh size	0.0546	0.0203	0.0391	0.0553
	(0.596)	(0.859)	(0.708)	(0.591)
breakfast days	0.412**	0.483**	0.392*	0.413**
	(0.007)	(0.006)	(0.013)	(0.007)
read days	-0.0927	-0.197	-0.125	-0.0924
	(0.342)	(0.051)	(0.208)	(0.343)
care age	0.0117	0.0652	0.0399	0.0117
	(0.896)	(0.519)	(0.661)	(0.896)
care age sq	-0.00035	-0.00120	-0.00063	-0.00035
	(0.719)	(0.274)	(0.527)	(0.720)
com-tab own	0.776	1.196*	0.687	0.776
	(0.080)	(0.016)	(0.123)	(0.080)
constant	200.9***	206.2***	204.5***	200.5***
	(0.000)	(0.000)	(0.000)	(0.000)
F-stat	46.561	40.908	44.715	46.581
No. Obs.	7489	5865	7202	7488

Note: Stars are based on p-values with * p<0.10, ** p<0.05, *** p<0.01 and clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics.

Table A.13: First-stage estimation results for log of school days

	(1)	(2)	(3)	(4)
provincial lock-down	-0.0469***	-0.0484***	-0.0457***	-0.0469***
	(0.000)	(0.000)	(0.000)	(0.000)
Covid cases (prv)	0.00046***	0.00045***	0.00047***	0.00046***
	(0.000)	(0.000)	(0.000)	(0.000)
std-tch ratio	0.00071	0.00085	0.00071	0.00071
	(0.113)	(0.064)	(0.121)	(0.112)
sch in urban	0.00759	0.00650	0.00867	0.00757
	(0.356)	(0.439)	(0.301)	(0.357)
private sch	-0.000479	-0.00585	-0.00152	-0.000468
	(0.968)	(0.611)	(0.901)	(0.969)
kindergarten	0.00264	0.00223	0.00273	0.00261
	(0.702)	(0.755)	(0.694)	(0.704)
frac of ECE	-0.0107	-0.0107	-0.0113	-0.0107
	(0.178)	(0.199)	(0.164)	(0.178)
frac of GOV	-0.00055	-0.00290	-0.00008	-0.00053
	(0.949)	(0.737)	(0.993)	(0.950)
frac of BA tch	0.0237	0.0224	0.0242	0.0237
	(0.159)	(0.185)	(0.149)	(0.159)
frac of MA tch	0.0195	0.0201	0.0199	0.0195
	(0.297)	(0.287)	(0.286)	(0.297)
Thai language	-0.0327***	-0.0315***	-0.0321***	-0.0327***
	(0.000)	(0.000)	(0.000)	(0.000)
Thai nationality	0.00509	0.00633	0.00392	0.00509
	(0.664)	(0.631)	(0.738)	(0.663)
care edu: M6	0.00223	0.00268	0.00147	0.00224
	(0.525)	(0.502)	(0.679)	(0.524)
care edu: Voc	-0.00511	-0.00355	-0.00637	-0.00510
	(0.192)	(0.410)	(0.112)	(0.193)
care edu: BA+	0.00590	0.00561	0.00660	0.00592
	(0.184)	(0.241)	(0.140)	(0.183)
wealth	-0.000865	-0.00205	-0.000469	-0.000876
	(0.624)	(0.273)	(0.795)	(0.619)
no parent	0.000276	0.00366	-0.00121	0.000330
	(0.943)	(0.352)	(0.760)	(0.932)
age	-0.177	-0.187	-0.191	-0.176
	(0.142)	(0.145)	(0.123)	(0.144)
age sq	0.0155	0.0165	0.0166	0.0155
	(0.106)	(0.105)	(0.093)	(0.107)
female	-0.00066	-0.00076	-0.00123	-0.00067
	(0.762)	(0.752)	(0.578)	(0.758)
special needs	0.00687	0.0105	0.00702	0.00704
	(0.180)	(0.120)	(0.172)	(0.171)
sick often	-0.00360	-0.00426	-0.00310	-0.00348
	(0.370)	(0.352)	(0.446)	(0.386)
weight	-0.00067*	-0.00064	-0.00071*	-0.00067*
	(0.043)	(0.056)	(0.036)	(0.043)
height	0.00057	0.00038	0.00056	0.00056
	(0.133)	(0.328)	(0.144)	(0.134)
hh size	0.000458	0.000202	0.000357	0.000462
	(0.527)	(0.801)	(0.628)	(0.524)
breakfast days	0.00276**	0.00338**	0.00263*	0.00276**
	(0.010)	(0.006)	(0.016)	(0.010)
read days	-0.00057	-0.00130	-0.00079	-0.00056
	(0.398)	(0.062)	(0.251)	(0.400)
care age	-0.00003	0.00037	0.00015	-0.00003
	(0.967)	(0.601)	(0.815)	(0.966)
care age sq	-0.00000105	-0.00000726	-0.00000280	-0.00000104
	(0.878)	(0.343)	(0.691)	(0.879)
com-tab own	0.00549	0.00858*	0.00486	0.00549
	(0.076)	(0.013)	(0.115)	(0.075)
constant	5.338***	5.398***	5.367***	5.335***
	(0.000)	(0.000)	(0.000)	(0.000)
F-stat	40.122	35.168	38.491	40.139
No. Obs.	7489	5865	7202	7488

Note: Stars are based on p-values with * p<0.10, ** p<0.05, *** p<0.01 and clustered at classroom level, F-statistics denote Kleibergen-Paap F-statistics.

F Ordinary Least Square Results

This section presents OLS estimates corresponding to the IV estimates in table 1-3 in section 4 of the main text.

Table A.14: Estimation results for the impact of school closure, no-school intensity, log of no-school intensity, school days and log of school days using OLS

	Cognitive Skills				Non-Cognitive Skills	
	Literacy	Receptive	Maths	Working Memory	Non-cog SDQ	Non-cog BPI
Panel A: School closure						
impact	-0.0388	-0.0678**	-0.0656**	-0.0088	0.0307	0.0407
RW p-value	[0.455]	[0.042]	[0.013]	[0.807]	[0.524]	[0.242]
p-value	(0.255)	(0.009)	(0.002)	(0.811)	(0.518)	(0.122)
No. Obs.	7489	7489	7489	5865	7202	7488
Panel B: No-school intensity						
impact	-0.918*	-1.186**	-0.777*	-1.248*	-0.427	0.0182
RW p-value	[0.067]	[0.033]	[0.059]	[0.060]	[0.788]	[0.969]
p-value	(0.052)	(0.006)	(0.021)	(0.025)	(0.519)	(0.966)
No. Obs.	7489	7489	7489	5865	7202	7488
Panel C: Log of no-school intensity						
impact	-0.532*	-0.690**	-0.445*	-0.687*	-0.272	-0.0025
RW p-value	[0.053]	[0.014]	[0.053]	[0.053]	[0.763]	[0.991]
p-value	(0.053)	(0.005)	(0.022)	(0.032)	(0.482)	(0.992)
No. Obs.	7489	7489	7489	5865	7202	7488
Panel D: School days						
impact	0.0018*	0.0024**	0.0022**	0.0027**	-0.0001	0.00002
RW p-value	[0.077]	[0.039]	[0.014]	[0.039]	[0.995]	[0.995]
p-value	(0.063)	(0.010)	(0.002)	(0.016)	(0.942)	(0.976)
No. Obs.	7489	7489	7489	5865	7202	7488
Panel E: Log of school days						
impact	0.249*	0.341**	0.313**	0.416**	-0.0132	-0.0109
RW p-value	[0.076]	[0.035]	[0.014]	[0.035]	[0.992]	[0.992]
p-value	(0.069)	(0.010)	(0.002)	(0.010)	(0.948)	(0.928)
No. Obs.	7489	7489	7489	5865	7202	7488

Note: RW p-values and p-values denote p-values with Romano-Wolf correction with 1,000 bootstrap replications and the traditional p-values both clustered at classroom level. The RW corrections were performed for each row at a time and for cognitive and non-cognitive skills separately. All estimations are clustered at classroom level. Stars are based on the RW p-values with * p<0.10, ** p<0.05, *** p<0.01.

G Calculation of Sampling Weight

This section describes how sampling weights are calculated. The weights account for the stratification procedure and non-responses of schools and students. There are six ingredients for sampling weight calculation.

1. Accounting for sampling of districts/amphoes: For each province except Phuket where there are only three districts, five districts (called amphoes in Thai) were randomly chosen by dividing all of them into five groups, $g = 1, 2, 3, 4, 5$, one is the central district (called amphoe Mueang in Thai), and the other four are ranked and equally divided using their poverty level. Let N_{pg} denote the number of districts in group g of province p. Therefore, the probability that a district d in group g of province p will be chosen is

$$P_{pd} = \frac{1}{N_{pg}} \tag{A.5}$$

Note that $P_{pd} = 1$ for the central district of every province p and all districts in Phuket since it has less than five districts.

2. Accounting for sampling of schools: Let N_{pds} and n_{pds} denote the total number of schools and the number of randomly chosen schools (both responded and non-responded) with size s (small, medium and large) in district d of province p. Therefore, the probability that a school j of size s in district d of province p will be chosen is

$$P_{pdj} = \frac{n_{pds(j)}}{N_{pds(j)}} \tag{A.6}$$

where $s(j)$ is the school size of school j. Note that since Phuket has only three districts, the survey employed a simple randomization without stratification. As a result, $P_{pd} = 1$ for all d of Phuket province, and the probability that a school of size s in district d of province p will be chosen is

$$P_{pdj} = \frac{n_{ps(j)}}{N_{ps(j)}} \tag{A.7}$$

where $N_{ps(j)}$ and n_{ps} denote the total number of schools and the number of randomly chosen schools with size s (small, medium and large) in Phuket province.

3. Accounting for sampling of classrooms: Let N_{pdj} and n_{pdj} denote the total number of classrooms and the number of randomly chosen classrooms in school j of district

d province p. Therefore, the probability that a classroom c in school j of district d province p will be chosen is

$$P_{pdjc} = \frac{n_{pdj}}{N_{pdj}} \tag{A.8}$$

4. Accounting for sampling of students: Let N_{pdjc} and n_{pdjc} denote the total number of students and the number of randomly chosen students in classroom c of school j in district d province p. Therefore, the probability that a student i in classroom c of school j in district d province p will be chosen is

$$P_{pdjci} = \frac{n_{pdjc}}{N_{pdjc}} \tag{A.9}$$

5. Accounting for non-response rates at the school and student levels: Note that not all randomly chosen schools responded in the survey. Let $\tilde{n}_{pds}$ denote the number of responded schools of size s in district d province p. Therefore, the response rate for school j of size s in district d of province p is

$$R_{pdj} = \frac{\tilde{n}_{pds(j)}}{n_{pds(j)}} \tag{A.10}$$

   Similarly, not all randomly chosen students were present at school on the survey date. We therefore have to account for non-responses. Let $\tilde{n}_{pdjc}$ denote the number of responded students in classroom c of school j in district d province p. Therefore, the response rate for student i in classroom c of school j in district d of province p is

$$R_{pdjci} = \frac{\tilde{n}_{pdjc}}{n_{pdjc}} \tag{A.11}$$

6. The last part is to account for different number of population in each province. Let N_p be the total number of kindergartners in province p.

Sampling weight for student i in classroom c of school j in district d of province p before adjusting for non-responses is

$$W^{br}_{pdjci} = \frac{1}{P_{pd}P_{pdj}P_{pdjc}P_{pdjci}} \tag{A.12}$$

Adjusting for the response rate gives the sampling weight as follows.

$$W^{ar}_{pdjci} = \frac{1}{R_{pdj}R_{pdjci}} W^{br}_{pdjci} \tag{A.13}$$

The final sampling weight can be achieved by adjusting for the population of each province as follows.

$$W_{pdjci} = \frac{W^{ar}_{pdjci}}{\sum_{i \in S_p} W^{ar}_{pdjci}} N_p \tag{A.14}$$

where S_p is the set of all responded/sampled children/students in province p.

References

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, volume 5, pages 111–150.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). Estimating the production function for human capital: results from a randomized controlled trial in Colombia. *American Economic Review*, 110(1):48–85.

Bentler, P. M. (1990a). Comparative fit indexes in structural models. *Psychological bulletin*, 107(2):238.

Bentler, P. M. (1990b). Fit indexes, lagrange multipliers, constraint changes and incomplete data in structural models. *Multivariate Behavioral Research*, 25(2):163–172.

Browne, M. W. and Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, 21(2):230–258.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Gorsuch, R. L. (1983). *Factor Analysis*. Psychology Press, New York.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate behavioral research*, 25(2):173–180.

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. Applied Psychological Measurement, Washington, DC.

Tucker, L. R. and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1):1–10.

UNICEF et al. (2017). *Overview: MELQO: Measuring Early Learning Quality and Outcomes*. UNESCO Publishing.