Studying during the pandemic: how young students were affected by the lockdown

Lucie NEIRAC (*), Laure HEIDMANN (*), Marianne FABRE (**)

(*) DEPP, Bureau des études statistiques et psychométriques sur les évaluations des élèves
(**) INSEE

`laure.heidmann@education.gouv.fr`

Mots-clés. : Crise sanitaire, éducation, analyse de données, doubles différences.

Domaines. Mesures et impact de la pandémie de Covid, Enseignement, éducation.

Résumé

En mars 2020, afin de lutter contre la pandémie de Covid19, toutes les écoles de France ont fermé leurs portes pendant au moins trois mois. Cette fermeture brutale a eu un impact profond sur la scolarité de millions d'élèves à travers le pays. Du jour au lendemain, les enseignants ont dû adapter leurs pratiques pour assurer une continuité pédagogique à distance, pour des élèves ne vivant pas toujours dans des conditions propices pour suivre une scolarité à la maison (familles nombreuses vivant dans de petites surfaces, sans connexion internet ni matériel adapté, etc). La question de l'impact de cette période sans précédent sur l'apprentissage des élèves apparaît alors cruciale. L'objectif de cette étude est d'y apporter de premiers éléments de réponse.

A partir des données des évaluations nationales exhaustives, qui ont lieu chaque année en France depuis septembre 2018 pour tous les élèves de CP et de CE1 (deux fois par an pour les élèves de CP, en septembre puis en janvier, et une seule fois en septembre pour les élèves de CE1), nous étudions l'impact de cette fermeture sur la progression des élèves qui l'ont vécue au cours de leur année de CP, la première de l'enseignement primaire. Nous calculons pour chaque élève un score agrégé par discipline (français et mathématiques), à partir d'une analyse en composantes principales des scores qu'ils ont obtenus dans les différents domaines évalués. Nous nous intéressons d'une part à la progression de ces scores (par discipline ou par domaine) entre la mi-CP et le début du CE1, ce qui nous donne une première mesure de l'effet du confinement, et cherchons d'autre part à observer si les conséquences qui en découlent sont uniformément réparties au sein de la population d'élèves étudiée.

Il apparaît sans surprise que la crise a eu un impact négatif sur la progression des élèves au niveau national. De plus, nous montrons que les élèves qui se trouvaient dans les positions les plus fragilisées avant la fermeture des écoles sont ceux qui en ont le plus souffert. Ainsi, les élèves scolarisés dans l'éducation prioritaire qui, lors d'une année normale, connaissent une progression moindre que leurs camarades du public, ont vu cette dernière chuter de manière plus importante suite au confinement, en français comme en mathématiques. C'est également le cas des élèves ayant débuté le CP avec des fragilités scolaires : dans les deux disciplines, moins bien armés que la moyenne face à la fermeture des écoles, cette dernière a pesé plus lourdement sur leur progression. Par ailleurs, nous mettons en évidence le fait que la crise n'a eu qu'un très

faible impact en compréhension orale du langage, domaine déjà très marqué socialement, où les compétences s'acquièrent principalement à la maison, beaucoup moins sur les bancs de l'école. Les conditions dans lesquelles les élèves abordent leur scolarité dans ce domaine sont déjà très inégalitaires, et l'école, même en temps normal, ne parvenant pas à les lisser, le confinement n'y a finalement pas changé grand chose. Il apparaît donc que la crise sanitaire a surtout creusé des inégalités sociales qui lui préexistaient.

Abstract

In March 2020, due to the Covid pandemic, all schools in France closed for at least three months. Using data from the exhaustive national assessments, we study the impact of this closure on the progression of pupils who experienced this crisis during their first year of primary school. We calculate an aggregate score per discipline (French and mathematics) from a principal component analysis of the scores obtained on the various assessment items, and then examine these scores using difference-in-difference models. We show that schools closure negatively impacts on the progression of all students while deepening pre-existing inequalities, and detail the factors by which it acted most strongly.

Contents

Remerciements

Nous souhaitons remercier l'équipe organisatrice dans son ensemble, et particulièrement Marc Christine, pour son travail de coordination ayant permis aux JMS de fêter leurs trente ans.

Nous tenons également à remercier Fabrice Murat pour l'aide précieuse qu'il nous a apportée. En débarrassant notre chemin de quelques os sur lesquels nous avions trébuché, et en apportant à notre travail la lumière de son expertise, il lui a permis d'arborer les couleurs qu'il revêt aujourd'hui.

1 Introduction

On March 17 2020, all schools in France had to close their doors to fight the spread of the COVID-19 pandemic. This unexpected closure had a profound impact on the schooling of millions of students across the country. From one day to the next, teachers had to adapt their practice to ensure continuity of their students' learning at a distance. Confined to their homes, students pursued their learning, sometimes with the help of parents and siblings, sometimes on their own. Three months later, when schools finally reopened and students were able to return to class, many report a decline in student achievement and a loss of classroom learning habits. Evidence is then needed to measure the impact of this unprecedented period on student learning.

To provide some basis for addressing this question, the national standardized assessments offer an extremely rich data set. Indeed, since September 2018, all children in France entering elementary school take three assessments over the course of first and second grades. The objective is to measure student progress in French and mathematics in their first year of schooling. These assessments are based on standardized tests, identical for all students in the country, taken at three points during the school year:

- the *early-first-grade* assessment that first graders take in September.
- the *mid-first-grade* assessment that first graders take in January.
- the *early-second-grade* assessment that second graders take in September.

The data collected on these national assessments track the progress of student learning throughout the first year of elementary school. In addition, it is possible to compare student results across years as nearly identical assessments have been administered every year since 2018. In particular, we are interested in the following two student cohorts:

- the cohort $\mathcal{C}_0$ with students who were in first grade in January 2019, the year before the pandemic. They experienced a normal schooling in first grade, and therefore constitute the *control* group.
- the cohort $\mathcal{C}_1$ with students who were in first grade in January 2020. They experienced school closures between the mid-first-grade and early-second-grade assessments, and constitute the *treatment* group.

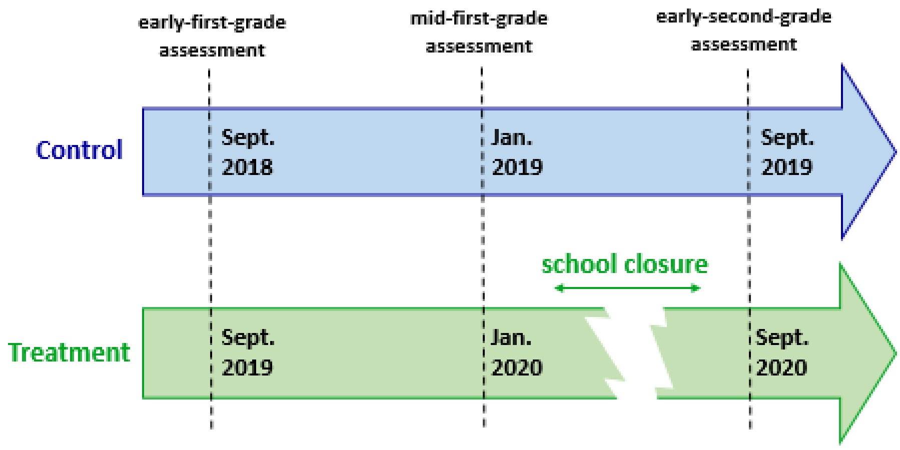


Figure 1: Diagram of the two student cohorts

Figure 1 provides a visualization of the data structure, with two cohorts and three assessments. Each assessment focuses on two disciplines, French and mathematics, which are divided into several skill areas – for example *Spoken word comprehension* in French or *Number recognition* in mathematics. Students' responses to a set of items in each domain allow to calculate scores per domain. Then, all domain scores are summarized into a composite score for the whole discipline.

As far as we know, such a dataset is almost unique. Engzell et al. [3] proposed a study on learning loss due to school closure during the pandemic, based on a dataset from bi-annual evaluations in the Netherlands, but those data are not exhaustive: they include only 15% of Dutch schools from which about 350,000 student aged 8 to 11 have been tested twice a year from 2017 to 2020. For our data, each cohort consists of almost all first graders in France, that represent about 800,000 children, who are assessed three times.

Once armed with this data, the first question that arises is whether school closures had a significant impact on students' academic outcomes. Which indicator should be chosen to describe student learning? Assessing the impact on student scores would provide some information, but it seems to us more interesting to look at the progression of student performance during this period. Indeed, what matters to us is not so much whether students were able to reach such and such a level at the end of the lockdown, but rather how the progression they should have experienced in a particular discipline or domain during a normal year was affected by the lockdown. Rather than performance, we focus on progress. Measuring the average impact of the lockdown on student progress is the first step. But is it reasonable to assume that all students have been affected in the same way? What factors played a decisive role in determining whether a student was more or less able to keep learning during the lockdown? These questions have been our guiding light during this study.

Descriptive statistics provide some initial answers. In order to capture more precisely the differential effect of school closure as a function of certain factors, we implement a difference-in-differences model in progression:

$$\Delta y_i = \beta_0 + \beta_1 T_i + \gamma^\top X_i + \nu^\top T_i X_i + u_i$$

In this model, Δy quantifies student academic progress between first grade and second grade, T indicates whether or not students are in the treatment group, X is a set of covariates according to which we seek to know whether the lockdown acted heterogeneously, and u represents unobserved or unaccounted-for variables in our model.

The analysis of the results shows that the lockdown had a negative impact on the progression of students between the middle of first grade and the beginning of second grade, and especially that this impact was not distributed homogeneously within the student population studied: the closure of the schools aggravated pre-existing inequalities.

2 Data

2.1 Assessments in French and mathematics

In France since 2018, all students entering elementary school take the same three assessments over the course of a year, in French and mathematics. The standardized tests take the form of exercise booklets that students complete by hand, following instructions given by their teachers according to a well defined protocol. Teachers then report their students' answers on a digital platform from which the results are centralized in the statistical service of the French Ministry of Education. The complete data consists of every student's response to every item of the test, as well as personal information about the student: school, gender and date of birth.

Students are evaluated in French and mathematics. The items that compose the assessments are grouped into skill areas. The assessments are almost identical between the two cohorts, but differ between the three points of schooling: early-first-grade, mid-first-grade and early-second-grade. In parallel with each evaluation, experiments are conducted to test new items on a sample of students. These items could then enrich the following years' assessments, either by replacing malfunctional items or by being added to the tests, which are therefore likely to be slightly modified from year to year. In order to get comparable data, only items which are common to both cohorts are taken into account in this study. In addition, speed exercises are excluded from the analysis because a change of item between the two cohorts can have an impact on the total score.

To get a more concrete idea of these tests, the assessments are briefly presented in the tables that can be found in Appendix A, summarizing the areas assessed and the number of items taken by each student cohort.

2.2 Calculation of student scores

2.2.1 Scores by domain of competence

Two types of scores are produced from the assessments to describe student performance: domain scores and discipline scores. Domain scores are defined as the sum of the points obtained by students on the items that compose the domain: each successful item contributes 1 point (0.5 point for short items) to the total domain score. For this study, domain scores are standardized across the two student cohorts.

Formally, we consider a student i who gets a score $s_i(j)$ at the item j. Let $S_i(d)$ be the raw score (i.e before standardisation) of the student on the domain $\mathcal{D}_d$, and $Y_i(d)$ the normalized score. The raw score is defined by:

$$S_i(d) = \sum_{j \in \mathcal{D}_d} s_i(j) \tag{1}$$

and the normalized score is obtained by:

$$Y_i(d) = \frac{S_i(d) - \bar{S}(d)}{\bar{\sigma}_S(d)} \tag{2}$$

where $\bar{S}(d)$ and $\bar{\sigma}_S(d)$ are respectively the empirical mean and standard deviation of $S_i(d)$ over the two cohorts.

To get an overview of the data, the following figures show students' raw scores in the domains that are common to the mid-first-grade and early-second-grade assessments. Scores are presented as histograms, distinguishing students by cohort. Note that the control group has significantly fewer students than the treatment group. This is due to a mismatch between assessment IDs and unique national student IDs that occurred during the 2018-2019 academic year and randomly impacted 13 % of the students in this cohort. An analysis is conducted to ensure the comparability of the two cohorts despite the missing student IDs in the first year, that can be found in Appendix B.

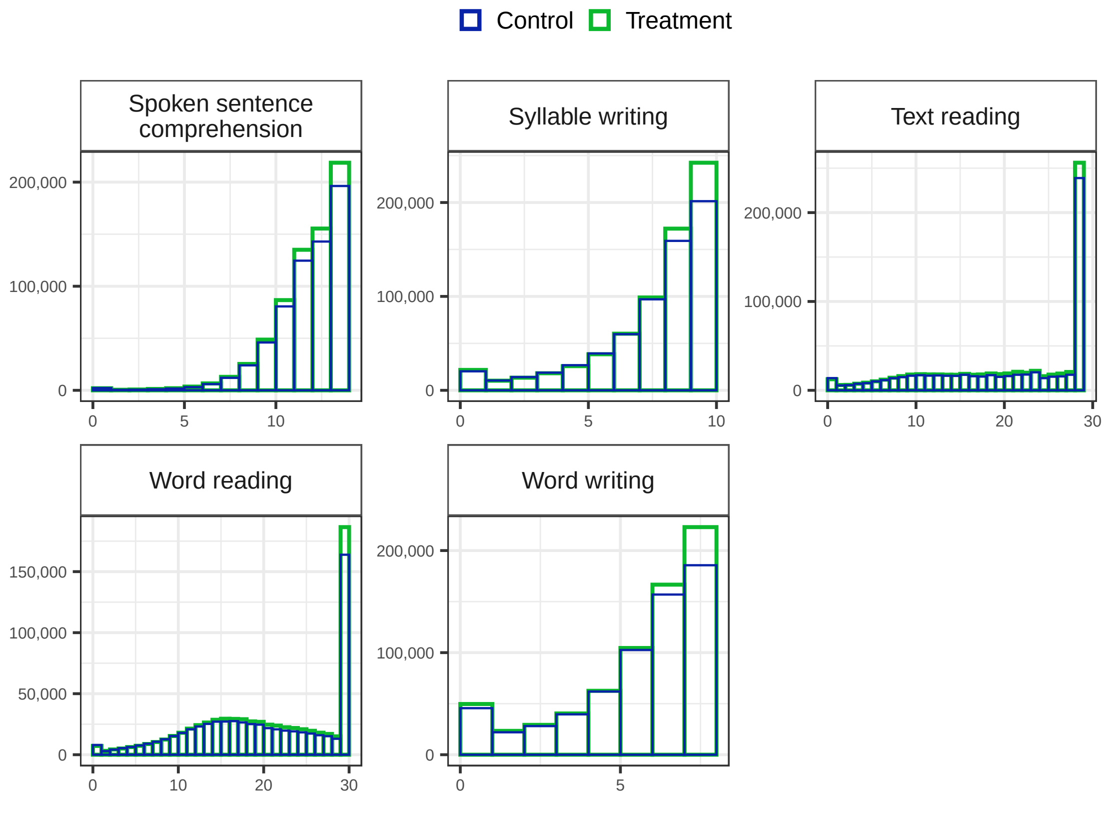


Figure 2: Raw domain scores at the mid-first-grade French assessment

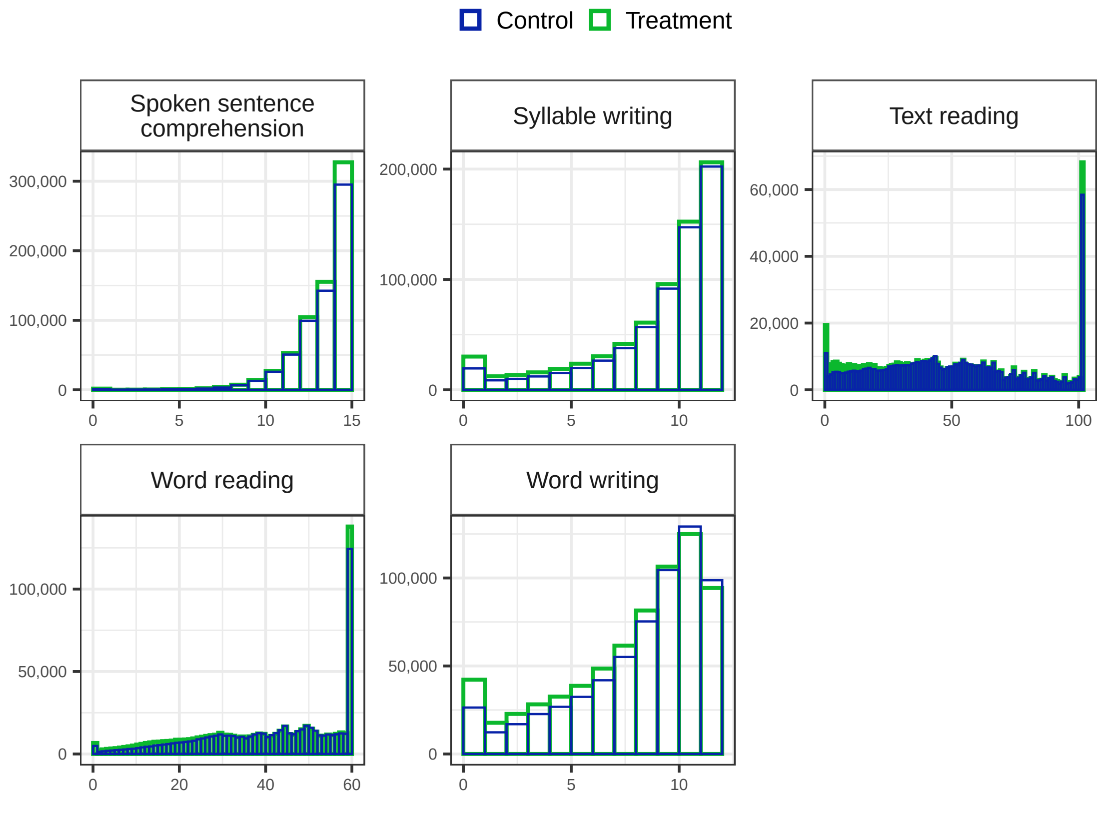


Figure 3: Raw domain scores at the early-second-grade French assessment

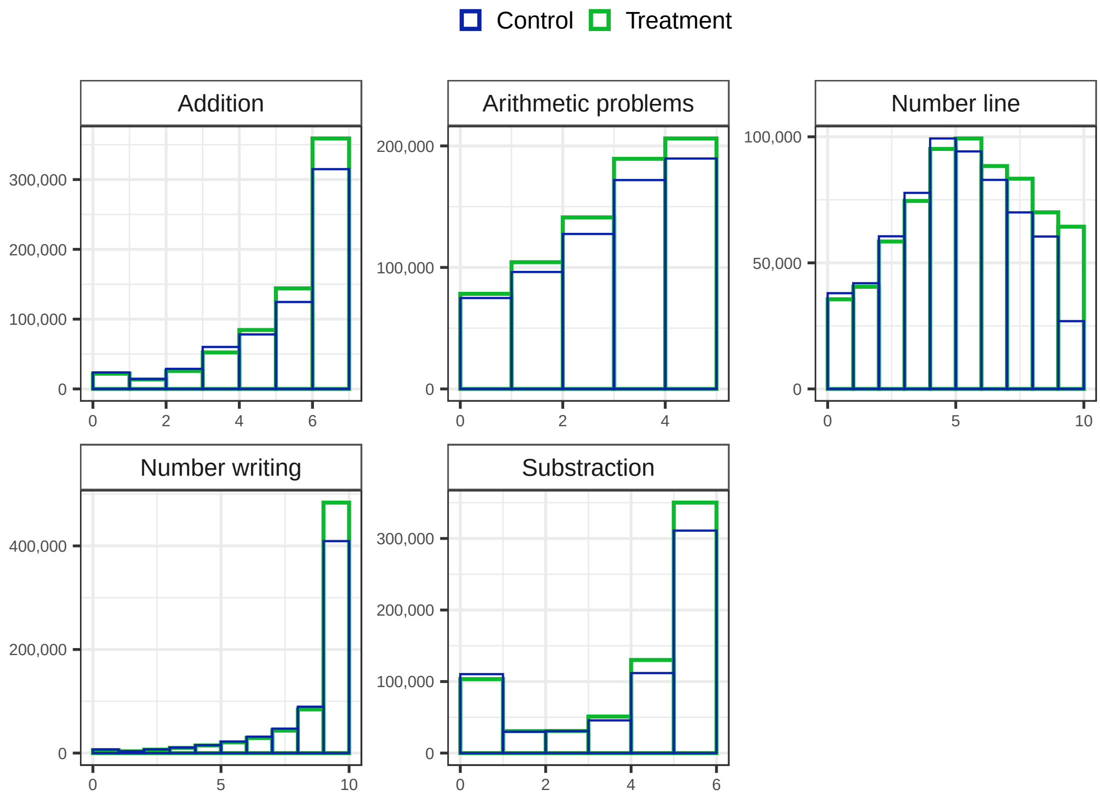


Figure 4: Raw domain scores at the mid-first-grade mathematical assessment

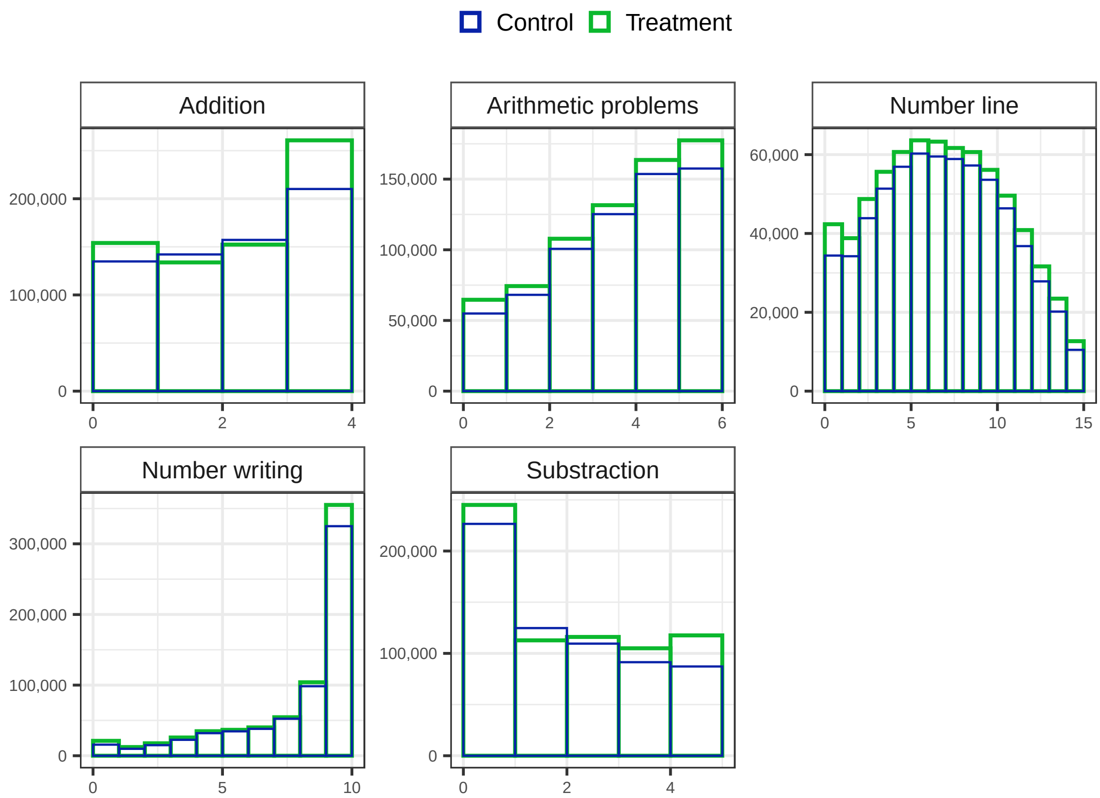


Figure 5: Raw domain scores at the early-second-grade mathematical assessment

2.2.2 Scores in French and mathematics

For educational purposes, domain scores are relevant to characterize students' abilities in different skills area. However, for the purposes of this study, discipline scores are necessary. Indeed, the objective is to compare the learning progression in first grade between students who had to leave school during the spring 2020 lockdown, and students who experienced a normal school year. Therefore, we need to build variables that measure student progress between the middle of first grade (before school closures) to the beginning of second grade (after school closures) for both cohorts. Domain scores provide a limited measure of student progress because few domains are common between the mid-first-grade and second-grade assessments, and of those domains, fewer items are usable (those passed by both cohorts). Consequently, a global score in the discipline, that would describe students performance to the whole assessment, provide more robust and comparable indicators for students progression in first grade.

To produce an overall score from the domain scores of a discipline, a principal component analysis is chosen since it can summarise the relative heterogenous information provided by each domain. Indeed, especially at this grade level, from early first grade to second grade, when children learn the foundational skills of reading, writing and counting, the exercises in the domains assess a broad and diverse range of skills: from deciphering letters to understanding sentences read by the teacher for instance. Nevertheless, it is possible to extract a single score that best captures students ability in the discipline from the domain results with a principal component analysis.

The analysis is performed on both cohorts together, at each time point of schooling separately. The data are composed of the domain scores for all students. There is a small proportion of students who have a missing score in at least one of the domains in an evaluation. This corresponds to the case where students were absent from class on the day of the domain assessment. To avoid removing completely these students from the data, their missing domain scores are imputed by their average score on the other domains of the assessment, once all scores are centered and reduced by domain. We get similar results in the PCA whether we impute missing scores or remove the affected students. More details can be found in Appendix C.

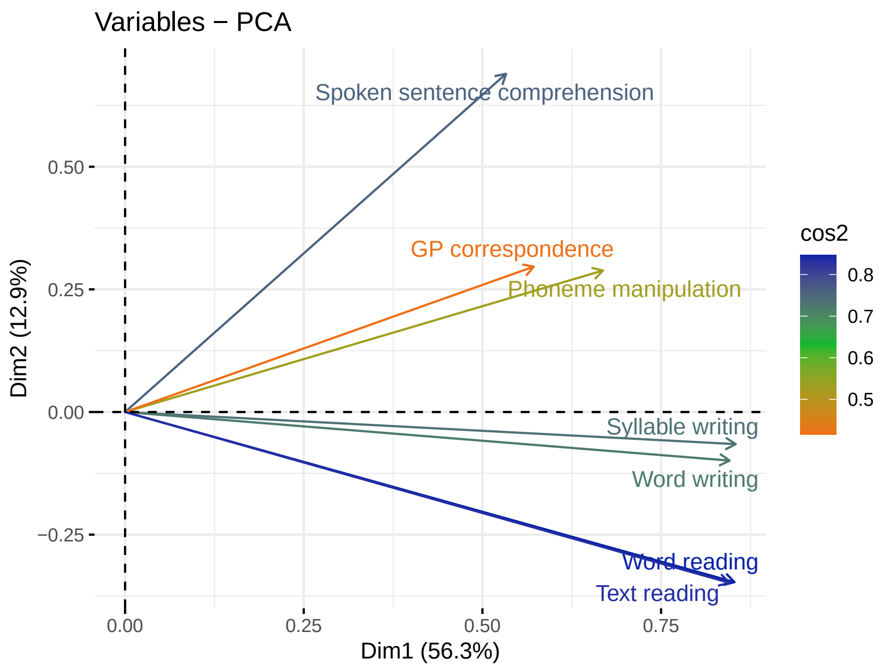


Figure 6: Projection of domains for the mid-first-grade French assessment

We run six PCAs: one on the domains assessing French and one on mathematics for the three data points early-first-grade, mid-first-grade and early-second-grade. Each time, the first dimension captures about more than 50 % of the variance. Figures (6) and (7) show the projection of the variables on the two first axes of the PCA for the mid-first-grade and early-second-grade French assessments. Similar figures on the others assessments are provided in Appendix C.

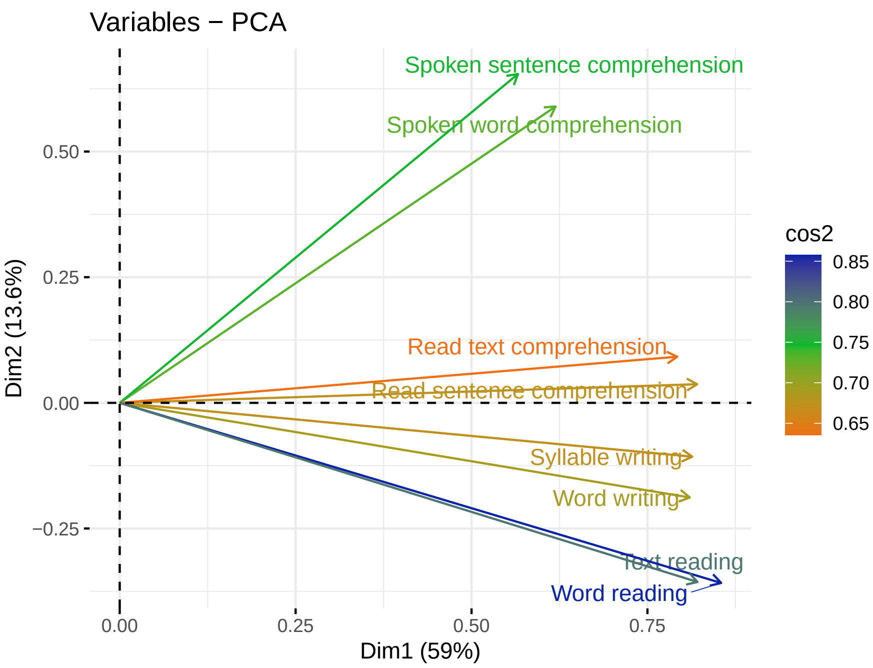


Figure 7: Projection of domains for the second-grade French assessment

The analysis produces, from a space where the observations are represented by domain scores, a new basis chosen so that the variance of the data is carried primarily by the first axis. The first component is then interpreted as students' overall ability in the discipline. The coordinates of the variables on the first component provide weights for the domains, so that we can compute a weighted average of the domain scores to build a composite score. This weighted average is equivalent to the projection of students onto the first axis, and summarise students' performance to the test.

Formally, the PCA computes the vector $(p_d)_d$ of the coordinates of domain variables on the first axis, so that we can construct the discipline score – or first coordinate – of student i from her vector of domain scores $(Y_i(d))_d$:

$$Y_i = \sum_{\text{domain } d} p_d Y_i(d) \tag{3}$$

The discipline scores are finally centered and reduced over the two cohorts.

2.3 Additional information on the students' profile

2.3.1 Learning context

Our data consisting of students scores are enriched by cross-referencing them with national databases to obtain additional information about the learning context.

In particular, the schooling sector is added, divided into three groups in the French context:

- private schools under contract (shortly named *private* hereafter)
- schools from priority education (*priority*)
- and the other public schools (*public*)

The priority sector gathers public schools that belong to a reinforced program because of the disadvantaged social profile of their students. More precisely, the priority education policy has been in place since 1981 to provide different support to public schools in disadvantaged context. These schools are categorized into two groups : the reinforced priority education network, which is made up of schools in isolated neighbourhoods or sectors with the greatest concentrations of social difficulties, and the priority education network, which concerns schools in more socially mixed neighbourhoods but with more significant social difficulties than those in the non-priority education network. In this study, we take the public (non-priority) sector as the baseline category.

In addition, the social environment of schools is captured by an index, called the social position index (SPI), which is constructed from information on the socio-professional categories of the parents of students enrolled in the school. It is a dimensionless standardized indicator available at the school level. The higher the SPI, the more favourable the family context is for students' academic success. Figure 8 shows the distribution of the SPI among students according to their schooling sector, separated by cohort. It can be seen that the priority sector has more students from disadvantaged backgrounds than the public (non-priority) sector, unlike the private sector, which has more privileged students.

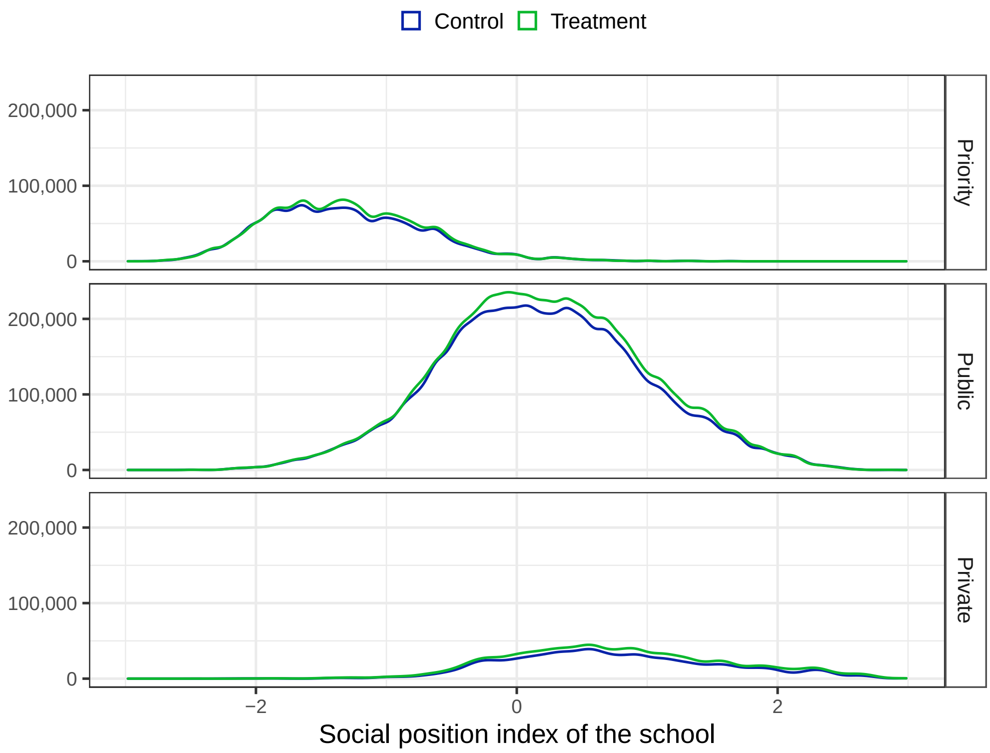


Figure 8: Distribution among students of the social position index

2.3.2 School closure in spring 2020

We are interested in measuring the effect of the school closure in spring 2020. Two indicators are available to describe the lockdown experiences of students. The first simply indicates whether students were in first grade during the spring 2020 lockdown (for the treatment group), or the year before (for the control group).

The second indicator adds the information whether students, who experienced school closures in 2020, went back to school before summer vacations or after. Indeed, the lockdown in France, which started in mid-March, lasted until mid-May, and it was up to parents to bring their children back to school during this period until the summer break. Therefore, some students did not return to school until the end of summer vacations. The impact of this additional lack of schooling must be measured. To that end, we collect responses to a questionnaire that was administered to all second graders during the September 2020 assessments. This questionnaire contained a variety of questions related to the students' experience of lockdown. In particular, one question, answered by the teacher, concerned the return to school before the summer vacations. Figure 9 shows the proportion of students who went back to school before the summer break, by schooling sector and by gender.

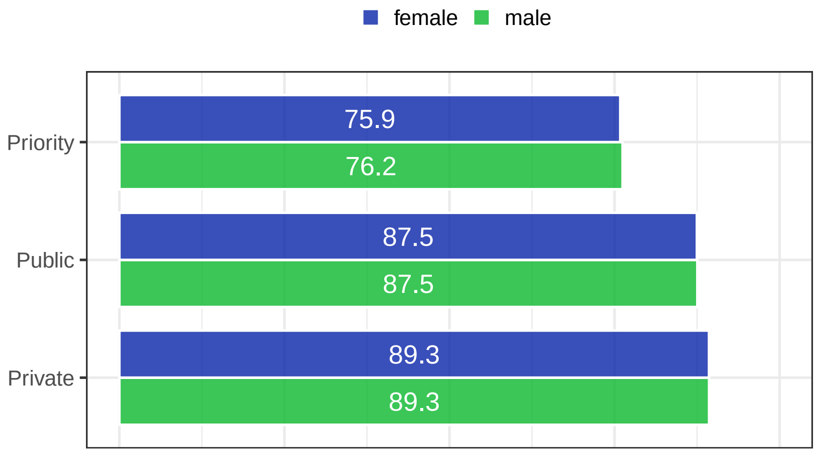


Figure 9: Proportion of first graders who returned to school before the summer break

This information allows us to construct a three-level treatment variable, distinguishing between students who were in first grade in 2019, those who were in first grade in 2020 and returned to school before the summer vacations, and those who did not return to school before the summer.

3 Methodology

3.1 Reminders on linear regression

We consider a variable of interest y (e.g. the student's score on a school assessment) and assume that it is generated by a set of p observed variables $x_k, k = 1..p$ (e.g. the student's sex or social background) and a set of unobserved variables (e.g. the conditions under which the assessment was given), whose effect is concentrated in a variable $u \in \mathbb{R}$, that is, there exists a function $f : \mathbb{R}^{p+1} \to \mathbb{R}$ such that $y = f(x_1, ..., x_k, u)$. We are interested here in the linear case, that is we assume there exists $\beta \in \mathbb{R}^{p+1}$ such that $f : (x_1, ..., x_k, u) \to \beta_0 + \sum_{k=1}^{p} \beta_k x_k + u$. Studying the variations of f by fixing u and all but one of the covariates x_{k_0} allows us to measure the causal impact of x_{k_0} on the outcome y. In order to estimate the parameters β_k from the observations of y and the p covariates for a set of N individuals, we classically use the least squares estimator, which is presented below.

3.1.1 Ordinary Least Squares estimator (OLS)

Let us denote $Y = (y_1, ..., y_N)^\top \in \mathbb{R}^N$, $\beta = (\beta_0, ..., \beta_p)^\top \in \mathbb{R}^{p+1}$ and let $X \in \mathbb{R}^{N,p+1}$ be the matrix obtained by concatenating the vector $\mathbb{1}_N := (1, ..., 1)^\top \in \mathbb{R}^N$ and the p columns-vectors $X_k \in \mathbb{R}^N$ made of the observations of each individual. The model to be estimated is the following:

$$Y = X\beta + u \tag{4}$$

Definition 1 (OLS estimator)**.** *In the previous framework, the Ordinary Least Squares estimator (OLS) is defined as the value of β which minimizes the euclidian norm of the residuals vector u:*

$$\hat{\beta}^{\text{OLS}} \in \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \ \|Y - X\beta\|_2^2 \tag{5}$$

3.1.2 Existence and properties of the OLS

We state here three classical assumptions that guarantee the existence of the OLS estimator and give it some convergence properties.

Assumption 1. *[Non-collinearity] There exists no collinearity relationship between the covariates $x_1, ..., x_p$. In other words, the matrix X is full-rank.*

Assumption 1 implies that the explanatory variables are chosen in such a way that they are not redundant, i.e. that they do not provide the same information several times in different ways.

Assumption 2. *[Exogeneity] The expectation of the residuals conditional on the covariates is zero:*

$$\mathbb{E}[u|X] = 0$$

Under Assumption 2, a change in the value of the covariates of a given individual does not lead to a change in the value of its residual. Moreover, it implies that the unobserved have a zero expectation, and are decorrelated from the observed covariates.

Assumption 3. *[Homoscedasticity] The variance of the residuals conditional on the covariates is proportional to the identity:*

$$\text{Var}(u|X) = \sigma^2 I_N \quad , \quad \sigma \in \mathbb{R}$$

Assumption 3 is a strong assumption which implies that the residuals of two individuals are not correlated. Moreover, we have that $\text{Var}(u_i|X_i)$ does not depend on the covariates, so that knowing the values of the x_{ik}'s for an individal i does not provide any information on the variance of its residual. Also, we have that $\text{Var}(u_i)$ does not depend on the individual i.

We are now ready to state the three following classical results about the OLS estimator.

Theorem 1. *[Existence and unicity] Under assumption 1, the OLS estimator (5) exists and is uniquely defined by:*

$$\hat{\beta}^{\text{OLS}} = (X^\top X)^{-1} X^\top Y$$

Theorem 2. *Under Assumptions 1 and 2, the OLS estimator (5) is unbiased, that is:*

$$\mathbb{E}[\hat{\beta}^{\text{OLS}}] = \beta$$

Theorem 3. *Under Assumptions 1, 2 and 3, the variance of the OLS estimator conditional to the covariates is:*

$$\text{Var}(\hat{\beta}^{\text{OLS}}) = \sigma^2 (X^\top X)^{-1}$$

Moreover, by denoting $\hat{u} = Y - X\hat{\beta}^{\text{OLS}}$, the parameter $\widehat{\sigma^2}$ defined below is an unbiased estimator of σ^2:

$$\widehat{\sigma^2} := \frac{\hat{u}^\top \hat{u}}{N - p - 1}$$

3.1.3 The difference-in-differences estimator

We consider a situation in which a population can receive a treatment or not, and where we have two time periods, t_0 and t_1: one part of the population (the control population) remains untreated in both periods, the other part (the treated population) receives treatment in the second period. We look at some outcome y for the whole population, and we are interested on the average treatment effect on the treated.
In our study, the population consists of children attending school in France who entered first grade in September 2018 or 2019. The first time period t_0 is the middle of first grade and the second period t_1 is the beginning of second grade. The treatment whose effect on students' schooling we want to assess is *having experienced a school closure in the second half of first grade during the March to May 2020 lockdown.* The treated (resp. control) population is the cohort $\mathcal{C}_1$ (resp $\mathcal{C}_0$) of children who entered first grade in September 2019 (resp. September 2018). The outcomes available to describe student learning are scores in French and mathematics, as well as detailed scores by domain.

Let us consider the following formal framework :

- $\tau_{i,t} = \mathbb{1}_{\{t=t_1\}}$ equals 1 if observation on individual i is made at period t_1
- $G_i = \mathbb{1}_{\{i \text{ belongs to the treatment group}\}}$
- $T_{i,t} = \mathbb{1}_{\{i \text{ has received treatment at time } t\}} = G_i \times \tau_{i,t}$
- $y^0_{i,t}$ is the potential outcome if no treatment
- $y^1_{i,t}$ is the potential outcome if treatment

At time t_0, only y^0 can be observed. At time t_1, we can measure the outcome y^1 for the treated population, whereas only y^0 is observed for the control population.
The objective is to identify the average effect of the treatment on the treated population:

$$\delta := \mathbb{E}[y^1 - y^0 | \tau = 1, G = 1]$$

.

This estimate can be decomposed into two parts:

$$\delta = \mathbb{E}[y^1 | \tau = 1, G = 1] - \mathbb{E}[y^0 | \tau = 1, G = 1]$$

The first component corresponds to the average outcome of the treated population after treatment – which can be measured at t_1 – and the second one is the average post-treatment outcome of the treated population if they had received no treatment. We do not have access to the second component since the population has been treated, however we do know the average post-treatment outcome of the control population : $\mathbb{E}[y^0 | \tau = 1, G = 0]$.

Therefore, the estimate can be written:

$$\delta = \big(\mathbb{E}[y^1 | \tau = 1, G = 1] - \mathbb{E}[y^0 | \tau = 1, G = 0]\big) - \big(\mathbb{E}[y^0 | \tau = 1, G = 1] - \mathbb{E}[y^0 | \tau = 1, G = 0]\big) \quad (6)$$

The first element in (6) is the difference in outcomes between the two groups at time t_1.
The second element corresponds to the difference between the observed post-treatment outcome of the control population and the potential post-treatment outcome of the treated population if they had not received the treatment. We do not observe this difference, nevertheless we can measure the difference in outcomes between these two populations before the treatment (at t_0) : $\mathbb{E}[y^0 | \tau = 0, G = 1] - \mathbb{E}[y^0 | \tau = 0, G = 0]$. If we can assume that, without treatment, the

difference in outcomes between the two groups would have been the same at t_0 and at t_1, we can measure every element to estimate δ. Therefore, we must formulate an hypothesis of a parallel evolution of the control group and the treatment group between times t_0 and t_1.

Assumption 4. *[Parallel Trend Assumption]*
In the absence of treatment, the average difference between treatment and control groups is constant over time:

$$\mathbb{E}[y^0|\tau = 1, G = 1] - \mathbb{E}[y^0|\tau = 0, G = 1] = \mathbb{E}[y^0|\tau = 1, G = 0] - \mathbb{E}[y^0|\tau = 0, G = 0]$$

There is no statistical test to control this key assumption. However, if we observe the outcome at several periods before treatment ($t = -1$, $t = -2$, ...), we can check the condition $\mathbb{E}[y^0|\tau = t, G = 1] - \mathbb{E}[y^0|\tau = t-1, G = 1] = \mathbb{E}[y^0|\tau = t, G = 0] - \mathbb{E}[y^0|\tau = t-1, G = 0]$ for any $t \leqslant 0$, since $y = y^0$ for those periods of time. Then, we can verify that y follows a parallel trend in the two groups before the advent of treatment, which makes it reasonable to assume that the hypothesis remains true after treatment.

Under Assumption (4), we can replace the second component of (6) by the difference between the groups at t_0 :

$$\delta = \big(\mathbb{E}[y^1|\tau = 1, G = 1] - \mathbb{E}[y^0|\tau = 1, G = 0]\big) - \big(\mathbb{E}[y^0|\tau = 0, G = 1] - \mathbb{E}[y^0|\tau = 0, G = 0]\big)$$

By permuting terms in parentheses, we notice that the first component is the difference in outcomes between t_1 and t_0 for the treatment group, and the second component is the equivalent for the control group :

$$\delta = \big(\mathbb{E}[y^1|\tau = 1, G = 1] - \mathbb{E}[y^0|\tau = 0, G = 1]\big) - \big(\mathbb{E}[y^0|\tau = 1, G = 0] - \mathbb{E}[y^0|\tau = 0, G = 0]\big)$$

In this form, we recognize the difference-in-differences estimator δ_{DiD}: the difference between the groups of the differences between the times. We are now ready to state the following theorem.

Theorem 4. *Under Assumption 4, δ can be obtained as the coefficient of T in the linear regression of y on G, τ and T.*

Proof. Let us consider the regression of y on G, τ and T:

$$y_{i,t} = \beta_0 + \beta_1 G_i + \beta_2 \tau_t + \beta_3 G_i \times \tau_t + u_{i,t}$$

(we recall that $T = G \times \tau$). We can easily check that Assumption (4) boils down to:

$$\mathbb{E}[u|\tau = 1, G = 1] - \mathbb{E}[u|\tau = 0, G = 1] = \mathbb{E}[u|\tau = 1, G = 0] - \mathbb{E}[u|\tau = 0, G = 0]$$

Now, under Assumption (4), a direct calculation leads to:

$$\begin{aligned}\delta_{DiD} &= \beta_3 + (\mathbb{E}[u|\tau = 1, G = 1] - \mathbb{E}[u|\tau = 0, G - 1]) - (\mathbb{E}[u|\tau = 1, G = 0] - \mathbb{E}[u|\tau = 0, G - 0]) \\ &= \beta_3\end{aligned}$$

which concludes the proof.

The difference-in-differences model is then the following:

$$y_{i,t} = \beta_0 + \beta_1 G_i + \beta_2 \tau_t + \beta_3 G_i \times \tau_t + u_{i,t} \tag{7}$$

and $\delta := \beta_3$ is what we call the difference-in-differences estimator of the average effect of the treatment on the treated.

3.2 Selected model specification

3.2.1 Progression model

The classical model to evaluate the difference-in-differences estimator was shown above in (7). In this specification, each individual appears two times: one time at t_0 and another time at t_1. It is possible to change specification so that it takes into account the panel structure of the data. The idea is then to analyse students' progression between t_0 and t_1 instead of their scores at these two times. The resulting model is called progression model and can be written:

$$\Delta y_i = y_{i,t_1} - y_{i,t_0} = \beta_2 + \beta_3 G_i + \Delta u_i \tag{8}$$

Then, δ is the coefficient of G in the linear regression of Δy on G.

In theory, models (7) and (8) give the same estimates for the coefficients β_2 and β_3. However, they may differ in the standard deviations of these estimates, since the progression model (8) incorporates the additional information that the observations y_{i,t_0} and y_{i,t_1} concern the same individual i. Although it is possible to add this information into the classical model (7) by clusterizing standard deviations at the individual level, we choose to use the progression model for its simplicity.

3.2.2 Treatment interaction on covariates

The progression model as specified in (8) allows us to estimate the average effect of the treatment on the population. However, in this study, we are interested in assessing the heterogeneity of this effect according to student characteristics. In particular, we would like to know how school closure affects students differently depending on their social background, and depending on their school performance at the beginning of the year for instance. Moreover, we wonder if the spring 2020 lockdown had the same effect on the academic learning of girls and boys.

In order to take into account the heterogeneity of the treatment according to certain characteristics, we can introduce some covariates to the previous model (8) and make them interact with the treatment. Formally, if we want to measure the effect of the treatment depending on a covariate x_k, the model to estimate becomes:

$$\Delta y_i = \beta_2 + \beta_3 G_i + \gamma_k x_k + \alpha_k G_i \times x_k + \Delta u_i \tag{9}$$

In this specification, the average effect of the treatment is still estimated by β_3. The coefficient γ_k represents the variation in Δy for the control group, generated by a one-unit increase in the covariate x_k if x_k is a continuous variable, or by the fact of being in the situation described by x_k – in comparison with the reference state – otherwise, all other things being equal. Moreover, the coefficient α_k corresponds to the interaction term between x_k and the treatment: it represents the additional variation in Δy on the treated generated by a variation in x_k. A non-significant α_k will mean that treatment does not impact differently the population according to the covariate x_k, while a significantly non-zero α_k will indicate a heterogeneity of the treatment across values of x_k.

In case of multiple covariates, let $X_i = (x_{1,i}, ..., x_{p,i})^\top \in \mathbb{R}^p$ be the vector of the covariates, the model (9) can be generalized:

$$\Delta y_i = \beta_2 + \beta_3 G_i + \gamma^\top X_i + \alpha^\top G_i X_i + \Delta u_i \tag{10}$$

3.2.3 Multilevel model

The models explained above assume that the observations are independent of each other. However, this assumption of independence is challenged as soon as the data are structured by levels, i.e. observations can be grouped within units and share common characteristics within the same unit, sometimes unobserved. This is typically the case for the data we use in this study: it is reasonable to expect that students in the same class share unobserved characteristics that influence the variable of interest, in this case the score in a domain or discipline. These characteristics may be related to the teacher (his or her experience, level of qualification), to the conditions under which the assessment is given, etc. Using a linear model while ignoring this particular structure of the data can lead to biased estimations of the standard errors. It is the purpose of multilevel models to take this dependency into account. In the literature, we find two types of statistical models that address the problems raised by the multilevel structure of a dataset: fixed effects models and random effects models. In this section we briefly outline the features of each of the two models, and explain why we have opted for a random effects model. Readers wishing to learn more about the statistical theory behind multilevel models can refer to [1] or [5].

In both cases, we start from the following regression model:

$$y_i = \beta_0 + \beta_1^\top X_i + \sum_{j=1}^{q} \left(\gamma^\top W_j + \alpha_j\right) \mathbb{1}_{\{i\sim j\}} + u_i \tag{11}$$

where y_i is an interest variable (e.g a student's score), X_i represents explanatory variables at individual level (e.g the gender), W_j represents explanatory variables at group level (e.g the scool sector), u_i (resp. α_j) represents unobserved (or unaccounted-for) variables at individual (resp. group) level, $\mathbb{1}_{\{i\sim j\}}$ indicated whether or not student i belongs to school j, and q is the number of groups. Exogeneity (Assumption 2) is always assumed for individual unobserved variables.

The fixed-effects model does not make any particular assumptions about group effects, which is considered as a whole (observed and unobserved variables). It boils down to putting $\nu_j = \gamma^\top W_j + \alpha_j$ in (11), which leads to the following classical linear regression model:

$$y_i = \beta_0 + \beta_1^\top X_i + \sum_{j=1}^{q} \nu_i \mathbb{1}_{\{i\sim j\}} + u_i \tag{12}$$

We can therefore estimate its parameters, those of interest being the betas. This model is particularly suitable when we want to take into account the effect of group variables, to refine the precision of our estimators, but without aiming at measuring it precisely.

The random effects model makes the – strong – assumption that unobserved group effects are independent of the explanatory variables in the model, that is: $\mathbb{E}[\alpha|X, W] = 0$ (in most cases, it will be assumed that $\alpha \sim N(0, \sigma^2)$). Equation (11) can then be rewritten as follows:

$$y_i = \beta_0 + \beta_1^\top X_i + \gamma^\top \sum_{j=1}^{q} W_j \mathbb{1}_{\{i\sim j\}} + w_i \tag{13}$$

where $w_i := \alpha_i \sum_{j=1}^{q} \mathbb{1}_{\{i\sim j\}} + u_i$ is an exogeneous noise. The parameters of this model can then be estimated, thus allowing, among other things, to measure the particular effect of the observed group variables.

For reasons of identifiability, it is not possible to include in a fixed effects model explanatory variables that are constant within each group. However, this is the case for half of the covariates we wish to introduce into our models in order to study the heterogeneity of treatment (school sector and SPI). This led us to discard the idea of a fixed effects model. We therefore had to make the assumption that our unobserved group variables, which include for example the conditions under which the assessment was taken, were independent of our explanatory variables, including the school sector and social position index. We are aware that this is a strong assumption that could be discussed.

3.3 Our model in practice

Our field of investigation encompasses two cohorts of students:

- cohort $\mathcal{C}_0$ with students who were in first grade in January 2019
- cohort $\mathcal{C}_1$ with students who were in first frade in January 2020

For each student in both cohorts, our dataset contains his or her score on the early-first-grade, mid-first-grade and early-second-grade assessments (by discipline or by domain). As explained in section 3.2.1, the panel structure of the data is taken into account by choosing a progression model rather than the classical difference-in-differences model. Therefore, we analyze differences in scores between early-second-grade and mid-first-grade in order to measure the effect of the lockdown on this progression. In particular, we are interested in measuring the heterogeneity of this effect according to a number of factors.

Furthermore, our data follow a hierarchical structure: students are grouped by school. It is more than reasonable to think that students in the same class and school share characteristics inherent to their schooling environment that significantly impact their academic performance. This is the case, for example, for the conditions in which the national assessment is taken. We therefore opt for a multilevel modeling and chose, for mainly computational reasons, to establish the groups at the school level. Multilevel models give two options: fixed effects models and random effects models. Since two of the variables for which we wish to test the heterogeneity in their interaction with treatment – namely schooling sector and social position index – are constant within groups, ruling out a fixed effects model, we evaluate a random effect model.

We examine the progression in score between early-second-grade and mid-first-grade, by discipline, French and mathematics, and by domain – for the common domains assessed at both measurement points. For this purpose, we implement the following regression models:

Classical linear regression:

$$\Delta y_i = \beta_0 + \beta_1 T_i + \gamma^\top X_i + \nu^\top T_i X_i + u_i \tag{14}$$

Random effects multilevel regression:

$$\Delta y_i = \beta_0 + \beta_1 T_i + \gamma^\top X_i + \nu^\top T_i X_i + w^\top Z_i + u_i \tag{15}$$

In the models (14) and (15),

- the outcome variable y is the score by discipline or by domain, Δy refers to the difference between early-second-grade and mid-first-grade scores.
- index i refers to students, with N the total number of individuals in the dataset involved in the regression.
- index j refers to schools, with q the total number of schools.

- T is the treatment variable whose effect is to be measured. Here, the treatment is belonging to the cohort that experienced school closure in the second part of first grade: $T_i = \mathbb{1}_{\{\text{first grade 2020}\}}$.
- $X_i \in \mathbb{R}^p$ is the vector of covariates that we choose to include in our model, which can be individual variables or group variables (i.e. constant within each group).
- $Z_i \in \mathbb{R}^q$ is a unit-vector with non-zero coefficient only at position j such that i belongs to school j
- $w \in \mathbb{R}^q$ is a random vector whose expectation is estimated to be zero conditional on the model variables
- u denotes the effect of unobserved (or not included in the model) individual variables, and is assumed to have zero expectation conditional on the other variables of the model.

4 Results

In this section, we present the main results of our study. We start with some descriptive statistics and then present the results of the experiments which we carried out by implementing the previous theoretical models with our data. All our experiments are conducted with R software.

4.1 Evolution of student performance between the two cohorts

To get a first look at the data, the evolution of student performance over the three assessments (early-first-grade, mid-first-grade and early-second-grade) is presented below, separated by cohort. The dots represent the average score obtained by students in each cohort on the three assessments. We recall that for each assessment, the scores are centered and reduced over the two cohorts. Consequently, the evolution of scores between the three assessment points does not measure the real improvement of the children's level. Indeed, the average evolution over the two cohorts is zero by construction, while students improve their skills in French and mathematics in first grade. The evolution of scores represents the progress of students in relation to their peers: a negative progression for a group does not imply that students are regressing but that they are progressing less than the other groups.

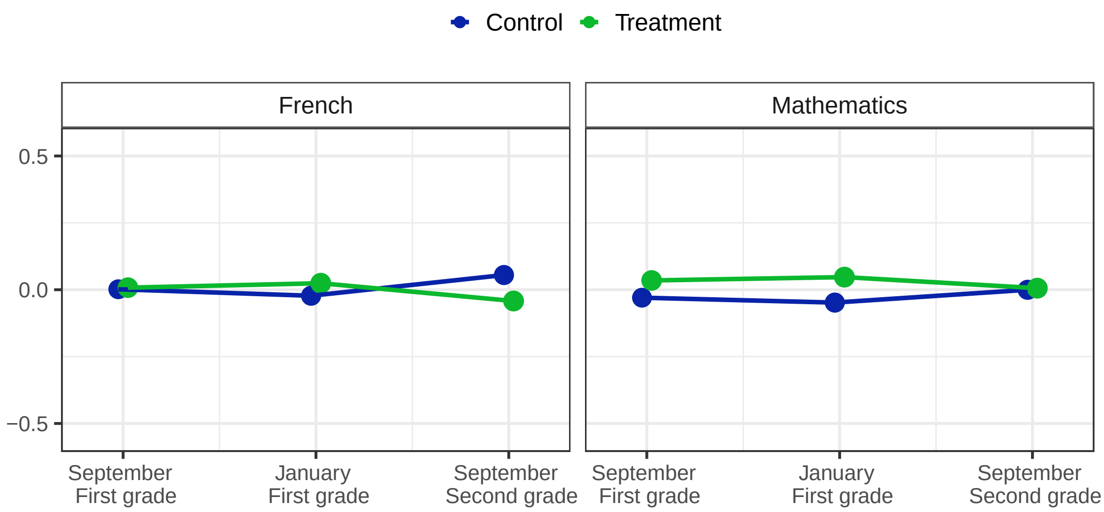


Figure 10: Average student scores in elementary school

In Figure 10, it can be noted that the two cohorts followed a parallel trend between the first two assessment times, in the first half of first grade, although the treatment population scored slightly higher, especially in mathematics. The effect of lockdown in the second part of the year can be observed in the crossing of the trends between the last two measurement points. The pupils in cohort $\mathcal{C}_1$, the treatment group, who were on average better than those in cohort $\mathcal{C}_0$ on the early-first-grade and mid-second-grade assessments, performed less in French and had about the same level in mathematics as those in cohort $\mathcal{C}_0$, after the school closure at the beginning of second grade.

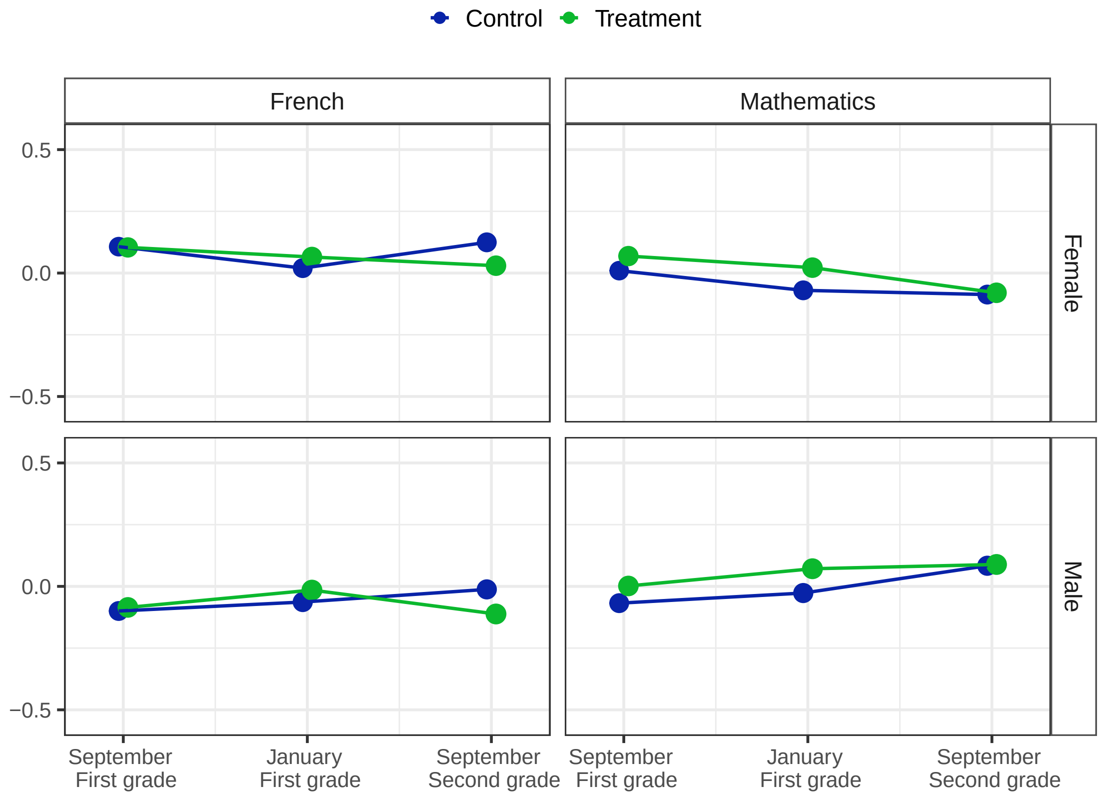


Figure 11: Average student scores by gender

When the average score is broken down by gender (Figure 11), parallel trends between control and treatment groups can still be observed in the first half of first grade. Moreover, in the second half, it can be noticed that student performance improves less, or declines more, in cohort $\mathcal{C}_1$ than in cohort $\mathcal{C}_0$. Regardless of the cohort, gender differences are measured: girls perform slightly better than boys in French, and slightly worse in mathematics, although the gap in mathematics, unlike the one in French, does not exist at the beginning of the first school year.

When analysing student scores by schooling sector (Figure 12), it is not surprising that the average score in both disciplines is lower in priority education than in public or private education. Furthermore, it appears that the effect of lockdown was particularly severe for pupils in priority education. Indeed, in this category, whether in mathematics or French, the average score shows almost no change between mid-first-grade and early-second-grade for cohort $\mathcal{C}_0$, whereas it fell for cohort $\mathcal{C}_1$. In the other sectors of schooling, the differences in the evolution of the average score seem much less significant between the two cohorts.

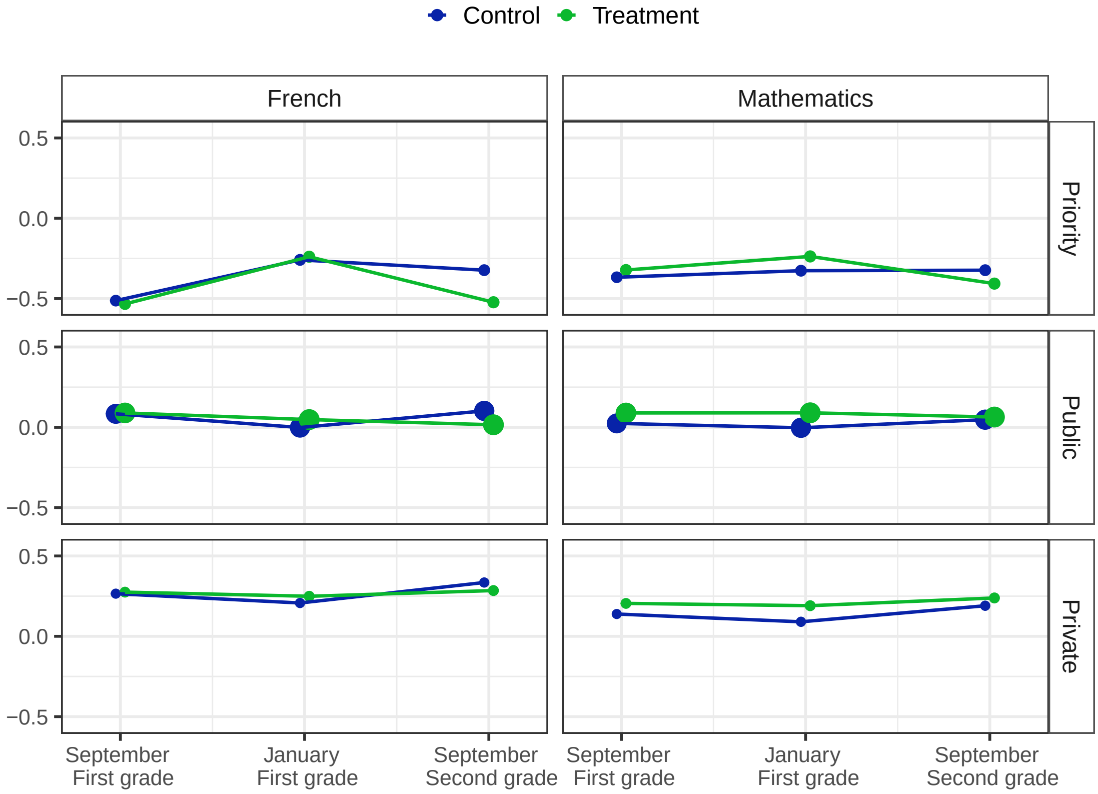


Figure 12: Average student scores by schooling sector

Similar figures detailed by skills areas can be found in the Appendix D.

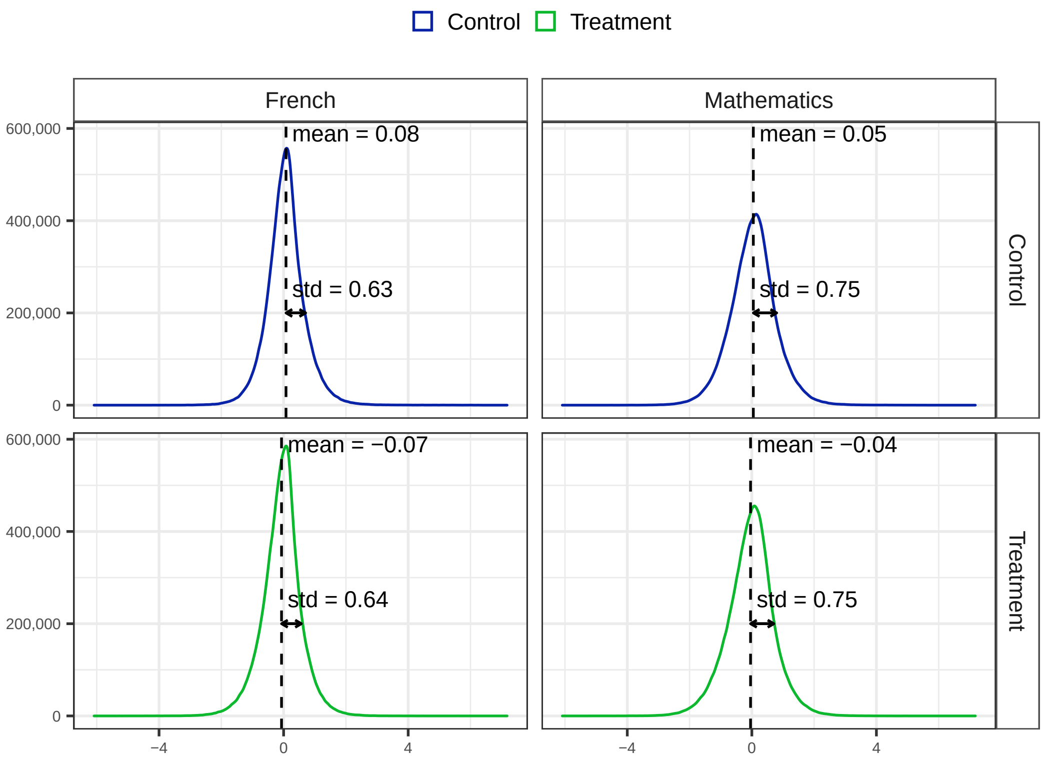


Figure 13: Distribution of student progress for both cohorts

In our study, we examine the variation of student progress between the two cohorts. Figure 13 shows the distribution of the difference in student scores between the early-second-grade and the mid-first-grade assessments, distinguishing students by cohorts.

4.2 Effects of school closures on learning French and mathematics

4.2.1 Measuring the impact of the lockdown

Here we study the impact of school closures on score progression from the mid-first-grade assessment to the early-second-grade. The reference group is the cohort $\mathcal{C}_0$ of students who were first graders the year before the COVID-19 pandemic. Having experienced the school closure in first grade is the treatment, written *year 2020* in the regression models, which is 1 for students in cohort $\mathcal{C}_1$, 0 for those in cohort $\mathcal{C}_0$.

Figure 14 show the results obtained by running model (15) respectively in French and mathematics. The estimated coefficients are represented with a 95 % confidence interval calculated from their standard errors. For ease of reading the graph, the coefficients are separated into two groups: those estimated from the control population, and the others measuring the additional effect on the treated population.

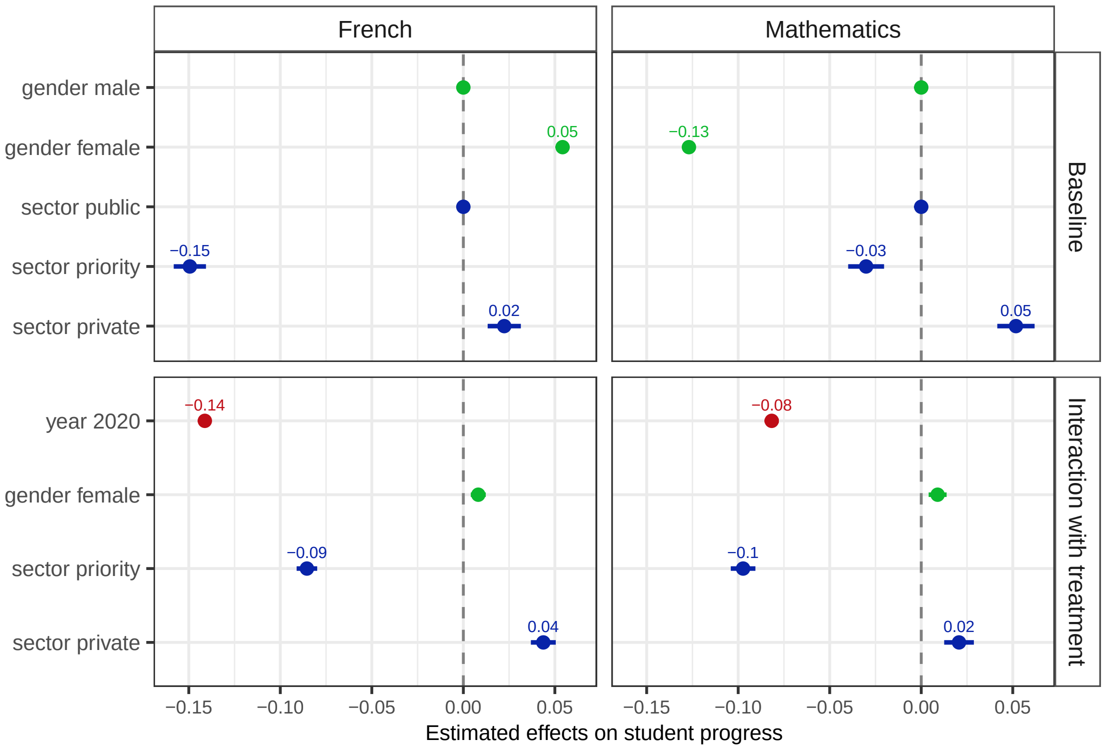


Figure 14: Effect of school closures on student progress

The regressions point out a significantly negative effect of school closure on the whole population: treated students experienced between mid-first-grade and early-second-grade a significant progression drop of −0.14 (22 % of the the standard deviation of the control population distribution) in French and of −0.08 (11 % standard deviation) in mathematics compared to the control population. Moreover, we can see that, compared to the reference student, who is a boy enrolled in a public (non-priority) school, the performance of girls, and to a greater extent of private school students, were less affected by the lockdown. On the other hand, students in priority education were highly more affected: we see that they experienced an additional drop in progress of -0.09 in French and -0.10 in mathematics (13 % standard deviation) attributable to school closure, compared to their peers from public education.

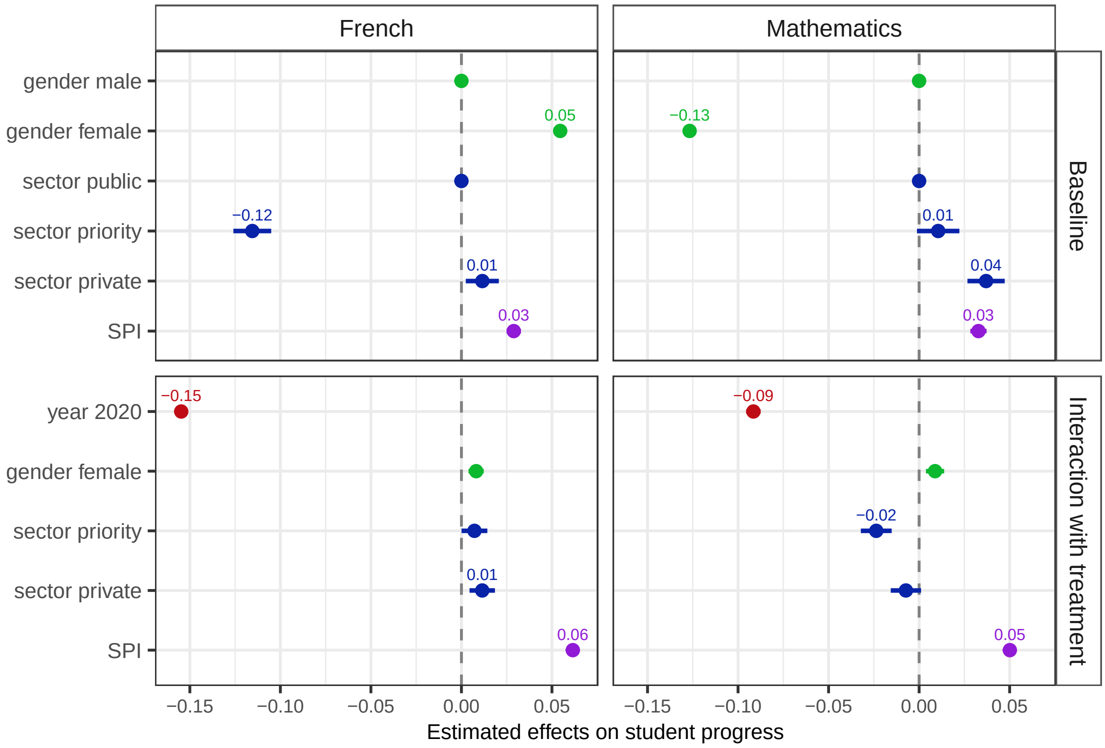


Figure 15: Effect of school closures including the social context

We then refine our model in Figure 15, including the social position index of the school (SPI) as a covariate. We then find that, in French as in mathematics, compared to the reference student, who is now a boy in public school with an average SPI, being a private school student during the lockdown has a much more mitigated effect on progression. So it seems that it is more their social background that benefits private sector students, than their school per se. On the other side, we observe that students from the priority sector experienced a lower decline in their French progression due to lockdown than students from the public sector. In mathematics, they still experience a larger drop in progress, but this effect is divided by five compared to the model that does not account for SPI (-0.10 to -0.02). All of this leads us to conclude that, at equivalent social context, priority education tended to benefit students who experienced school closures.

Finally, in Figure 16, we add a last covariate to our model: the initial score in the early-first-grade assessment. First of all, we observe that the effect of this variable on progression is negative, both in French and in mathematics, respectively -0.04 and -0.03. This indicates that students starting first grade with a high level have a narrower margin of progression than those starting from a lower level. This effect also reflects the type of academic skills that are at stakes in first grade: reading, counting and writing. More than in subsequent grades, these foundational skills constitute a sort of threshold to cross, so that students entering first grade with very limited reading skills, for example, will make considerable progress once they are able to read. However, students with a higher starting level were better equipped to deal with school closures: the decline in progress due to the lockdown was reduced by 8 % standard deviation in French, and 3 % in mathematics, for students who scored one standard deviation higher than the initial average level.

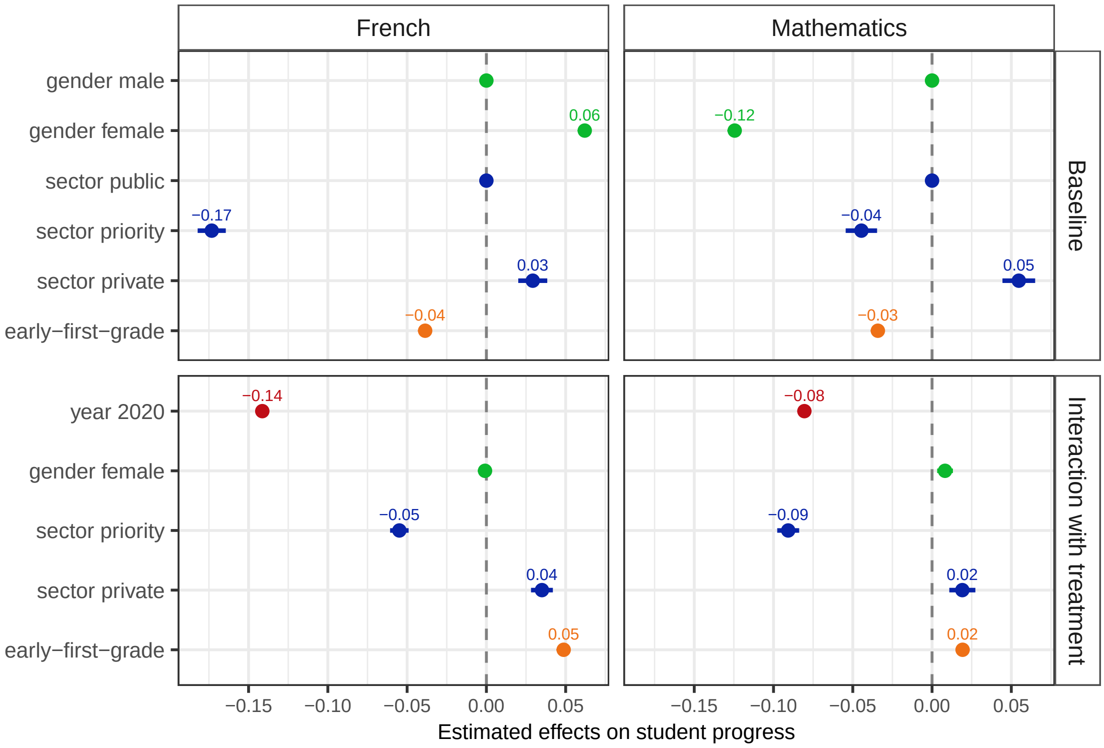


Figure 16: Effect of school closures including students initial performance

We also implement an ordinary least square model, as defined in the model (14), with the same sets of control variables. The results obtained are presented in Tables 7 and 8 in the Appendix E. The conclusions we draw are the same as with the previous model, but we obtain slightly lower standard errors on estimated coefficients compared to the multilevel model.

4.2.2 Heterogeneity of effects by social environment

The previous analysis revealed heterogeneity in the effect of school closures along the social dimensions of our study: school sector and social position index. These two dimensions are correlated since, on the one hand, the priority education network groups together schools in disadvantaged neighborhoods and, on the other hand, private schools tend to attract more advantaged students. In order to more accurately separate the effect of social context and school sector, we create a new typology of schools. The idea is to compare the priority and private sectors with public schools that have students from similar social background. To do so, we separate the public (non priority) sector into three groups : *public disadvantaged* with the 15 % of students with the lowest SPI, *public advantaged* with the 30 % of students with the highest SPI, and *public average* with the remaining public sector. Figure 17 shows the distribution among students of the school's social position index for the five groups in this new topology.

Schools in the priority education network have student from disadvantaged background but benefit from certain remedial measures. For example, since 2017, a policy has been in place to reduce the size of first and second grade classes in priority education schools, aiming for 12 students per class, instead of 24, which represents a near doubling of the number of classes compared to the situation before the measure. In contrast, schools in the *public disadvantaged* group have students from social categories almost comparable to those in priority education but do not benefit from any remedial measures. We run model (15) with this new school sector typology as covariates to measure whether specific instructional protocols in priority and private education impact student progress during the lockdown, independent of social context. Results are presented in Figure 18.

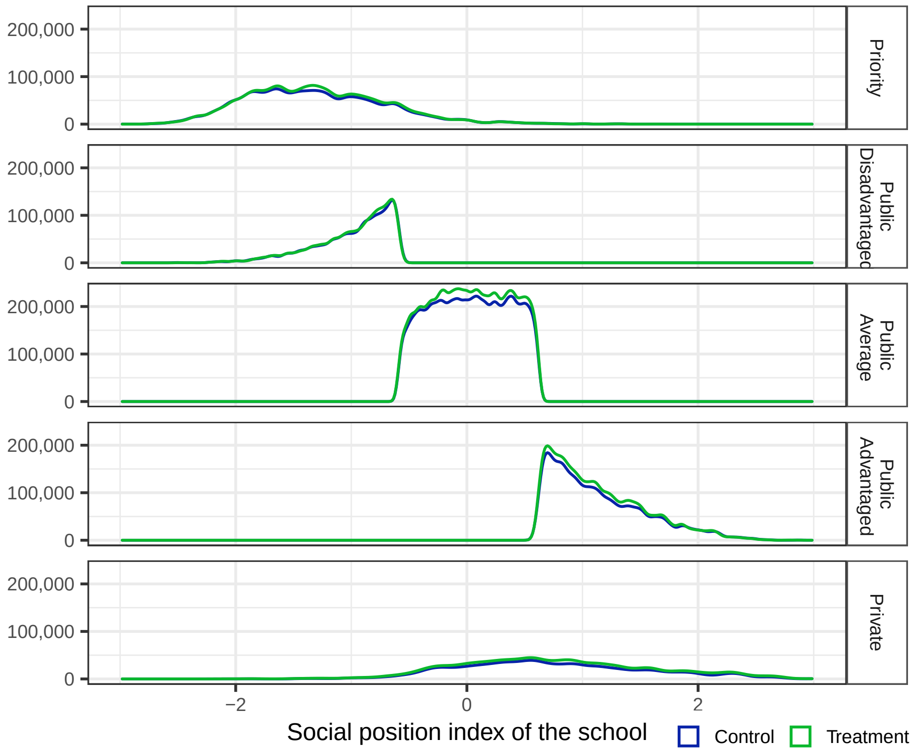


Figure 17: Distribution of social position index by alternative school typology

In French, in normal conditions, both priority and disadvantaged public schools students make less progress between mid-first-grade and early-second-grade than the average pupil (a boy in an average-SPI public school), but this progress deficit is eight times greater for priority education pupils. It shows that the two categories are not completely comparable since the most disadvantaged schools are in the priority network (as it can be seen in Figure 17). When we look at students who experienced lockdown during first grade, this trend is maintained, but the gap between these two groups is reduced: the additional drop in progress in priority schools and disadvantaged public schools are almost similar.

The same pattern emerges in mathematics : priority and low-SPI public schools pupils always make less progress than the average pupil, and this additional decline is 3 times higher for priority education pupils in cohort $\mathcal{C}_0$, while it is only 1.5 times higher when looking at the cohort $\mathcal{C}_1$. This suggests that the remedial measures in place in the priority education schools continued to have an effect even during the lockdown. It is reasonable to think, for example, that it is easier for a teacher to carry out quality remote monitoring when class size is reduced. For students from the disadvantaged public sector, the sharp drop in progress observed directly reflects the difficulty of disadvantaged families to manage this period.

At the other end of the social scale, students in private schools and in public schools with high SPI make more progress than the average student between the middle of first grade and the beginning of second grade. On average, students in advantaged public schools have a higher SPI than students in private schools (Figure 17), their progress in French is consequently slightly higher, but this is not the case in mathematics, which may indicate that the private sector is more successful in mathematics. After the lockdown, the decline in progress of these students due to the school closure was significantly lower than that of the average student, and more pronounced for advantaged public sector students. This shows that social context plays a more important role in learning during the school closure than does being in the private or public sector.

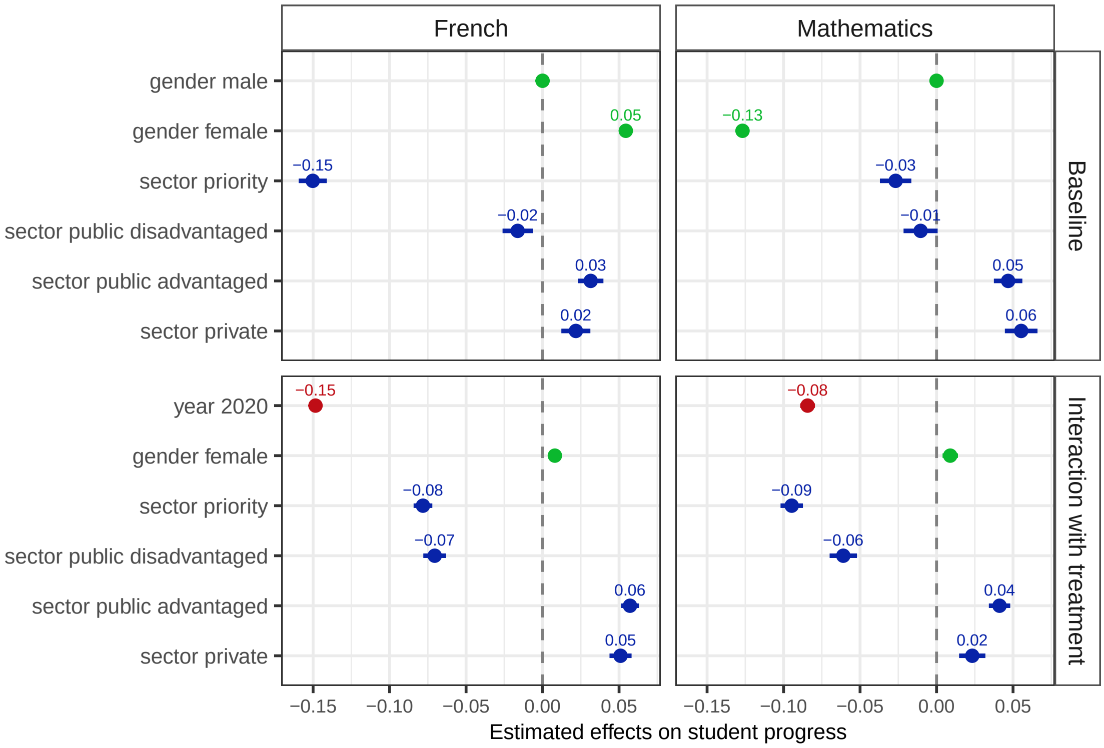


Figure 18: Effect of school closures with an alternative school categorization

In particular, the gap in progress between students from the advantaged public sector and the average student is three times greater in cohort $\mathcal{C}_1$ than in cohort $\mathcal{C}_0$ in French (5 % standard deviation to 15 % standard deviation) and doubles in mathematics (7 % standard deviation to 12 % standard deviation). This indicates that the correlation between social background and learning progress is exacerbated by the school closure, during which students from more privileged social backgrounds were given more incentives to engage in activities such as reading that could boost their performance especially in French, resulting in a widening of the gap between rich and poor students.

4.2.3 Difference between students who returned to school before summer and those who did not

In order to study the impact of returning to school or not before the summer break, we use a three-level treatment variable that was presented in section 2.3.2. The results are summarized in Figures 19 and 20. In the figures, the coefficients are separated into three groups: those estimated from the control population, the others measuring the additional effect on the treated population who returned to school before summer vacations, and the last ones on the treated population who returned to school after summer vacations.

We observe that the students in cohort $\mathcal{C}_1$ who did not return to school experienced a drop in their progression of about twice as much in French and three times as much in mathematics as their classmates who returned to school. This extra time without school seems to have particularly affected female students. Indeed, while girls who returned to school before the summer suffered less from the lockdown than their male peers (their decline in progression was less than that of boys), those who did not have this opportunity experienced a greater drop in progression than their male classmates, and this can be observed in French as well as in mathematics.

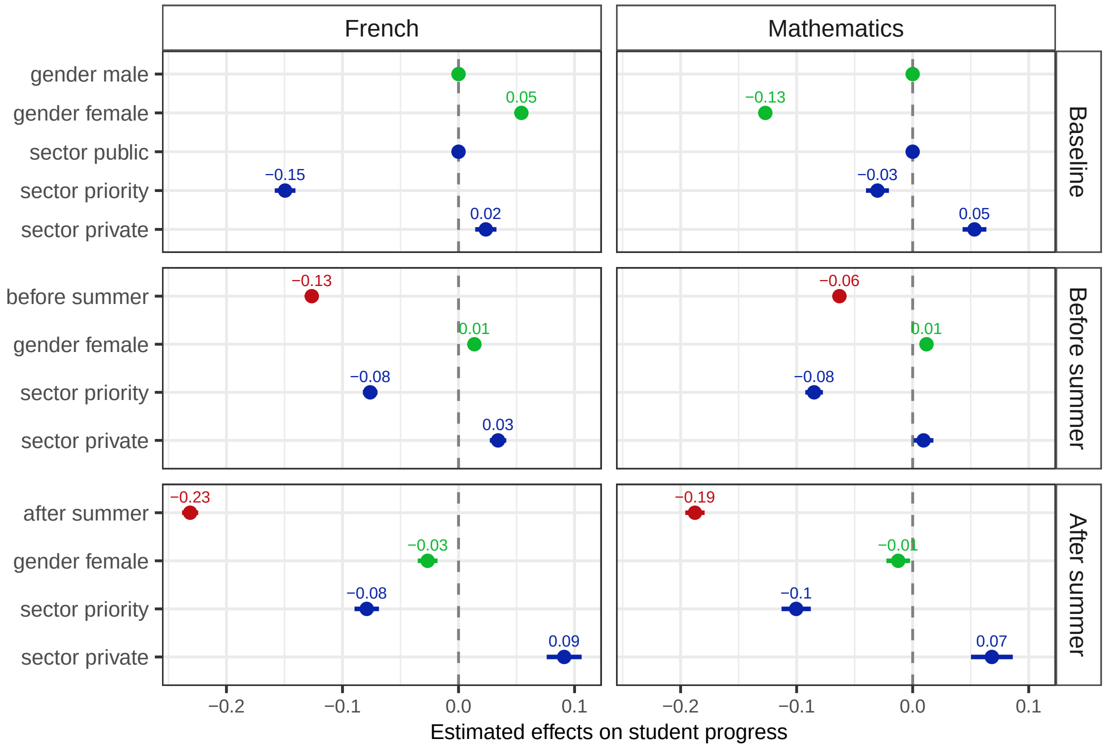


Figure 19: Effect of returning or not to school before summer break

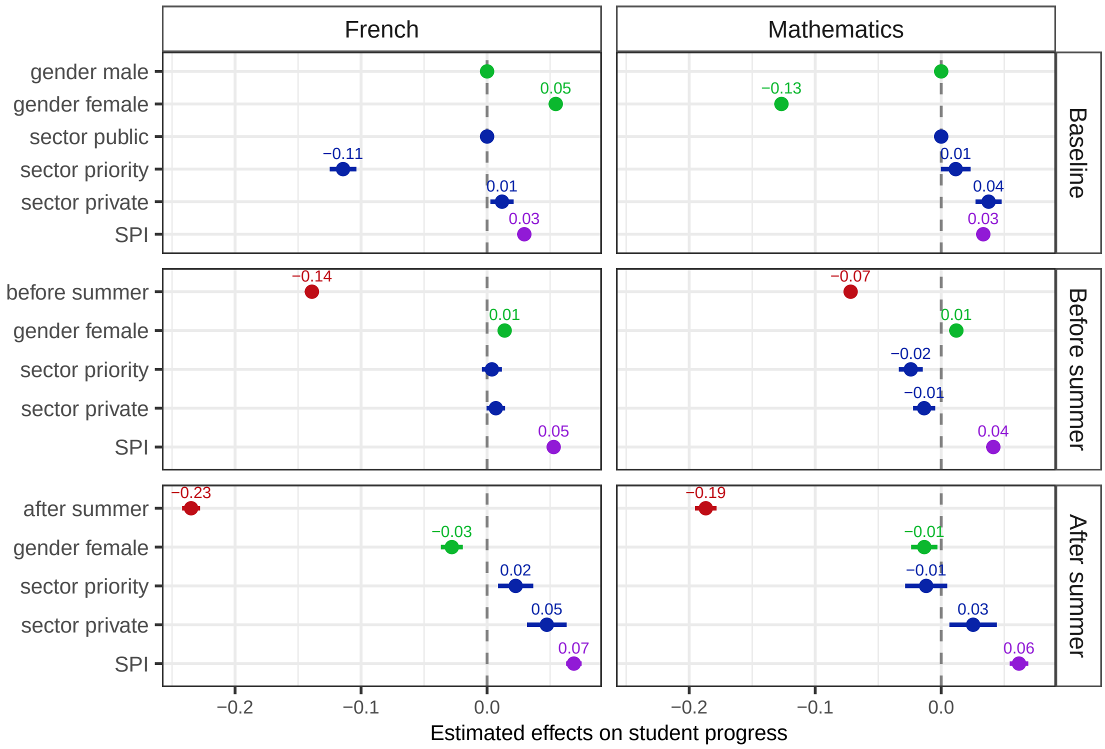


Figure 20: Effect of returning or not to school including the social context

When using school sector as the sole social determinant (Figure 19), being educated in a private school has a significantly positive effect among confined students, whether or not they returned to school before the summer. However, this effect is much greater for students who did not return to school, in both disciplines. When the SPI is introduced in addition to the school sector as a socially determined variable (Figure 20), the effect of private education becomes non-significant in French for students who have returned to school. Among those who have not returned to school, it keeps a significantly positive effect, but it is half as large as when the SPI is not taken into account. In mathematics, this effect is further amplified: when the SPI is introduced, among the students who have not returned to school, those in private education continue to experience a significantly smaller decrease in progression than their peers in public education (with a smaller gap than in the case of the model without the SPI), while among the students who have returned to school, those in private education see their progression drop significantly more than their mates in public education. For an equivalent SPI, and as far as mathematics is concerned, the return to school was therefore more beneficial if the school was public rather than private. All this corroborates the observation made earlier: it is more the privileged social environment in which they evolve than the school structure as such that favors private school students.

With regard to priority education, when the SPI is not taken into account (Figure 19), its effect is significantly negative for the pupils in cohort $\mathcal{C}_1$, whether or not they have returned to school, with an increased effect for those who have not returned to school, in both disciplines. On the other hand, as soon as the SPI is introduced (Figure 20), for equivalent SPI, students who had experienced school closures had a similar or smaller drop in progression in French if they were enrolled in priority education than if they were enrolled in public schools. In mathematics, priority education retains a significantly negative effect, both for pupils who had returned to school and for those who had not, but this effect is much smaller than the case where the SPI is not taken into account.

A high SPI has a positive effect on score progression for pupils who experienced the lockdown in first grade, and this effect is stronger for pupils who did not return to school before the summer break, which can be easily explained by the fact that pupils whose parents belong to the most affluent socio-professional categories benefited from better learning conditions during the lockdown, and therefore suffered less from the closure when it extended until summer.

4.3 Effects on student progress by domain

To better understand the impact of school closures on specific skills, we examine the difference in scores by domain between the two cohorts. Among the domains evaluated in French and mathematics, some are common to the mid-first-grade and the early-second-grade assessments. On these domains, it is therefore possible to measure the impact of school closure on students progress using our methodology. In order to conduct our analyses, we standardise these scores by domain over the two cohorts of students. Details on the common domains evaluated are provided in Appendix A.

To get a sense of the social determinant in learning specific skills, we present, in Figures 21 and 22, the relationship between students' scores and their school's social position index for each domain on the two assessments studied, comparing the two cohorts. First, a strong link between good academic performance and a privileged social background is blatant for all domains at both grade levels. It is also noticeable that the trends are parallel between the two cohorts at the mid-first-grade assessment (before the lockdown), confirming that the two groups are comparable before the school closure. However, at the second-grade assessment, most of the slopes become steeper for the treated group: the importance of social context on learning progress is exacerbated after the lockdown.

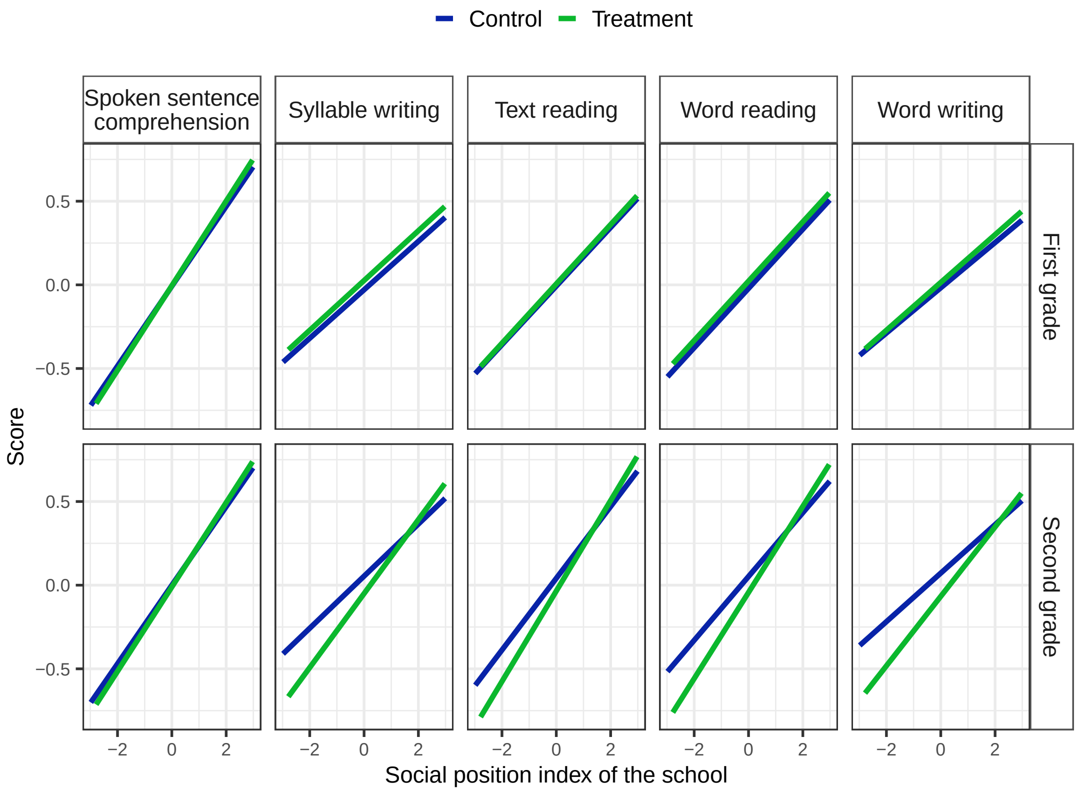


Figure 21: Influence of social context on student performance in French

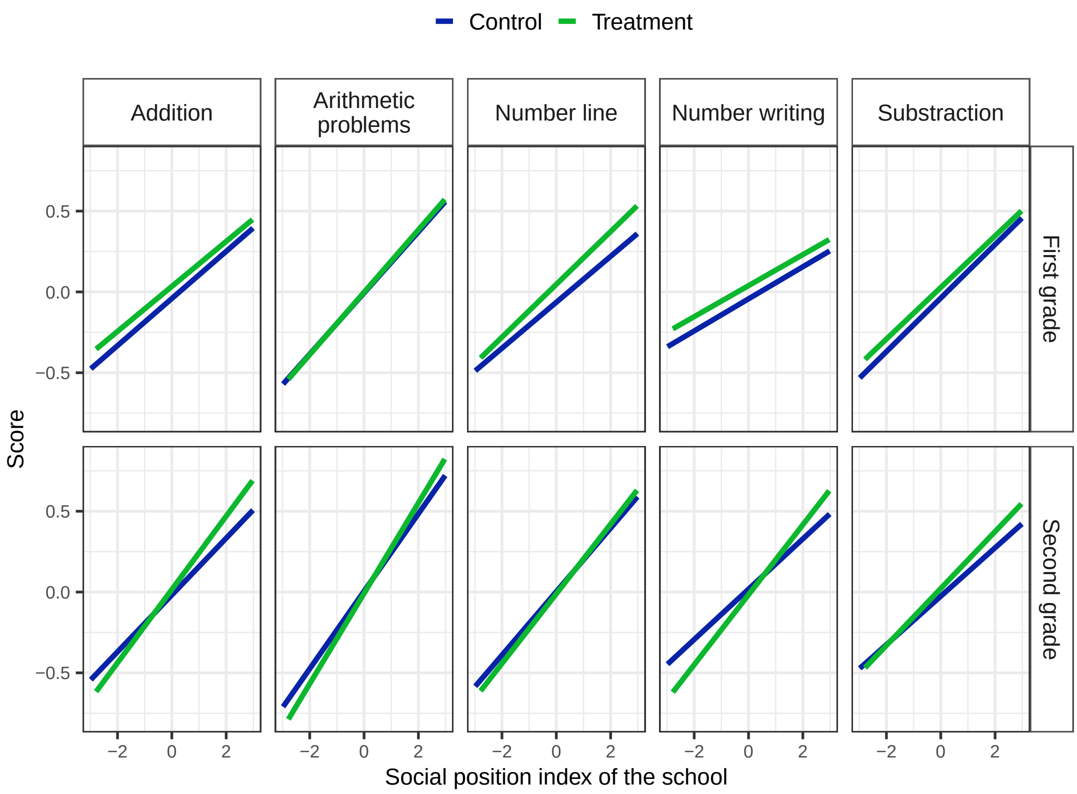


Figure 22: Influence of social context on student performance in mathematics

Domain scores constitute less robust indicators of student performance than discipline scores since they are based on a smaller amount of items. Therefore, we investigate the impact of school closure on student progress by domain, without measuring heterogeneity of effects. For this purpose, we implement multilevel regressions, described in the model (15), and the results are presented in Tables 1 and 2.

Table 1: Effect on student progress by domain in French

	Dependent variable:				
	Difference in scores between early-second-grade and mid-first-grade				
	Spoken sentence comprehension	Syllable writing	Text reading	Word reading	Word writing
	(1)	(2)	(3)	(4)	(5)
year 2020	−0.013***	−0.159***	−0.092***	−0.140***	−0.171***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Constant	0.006***	0.085***	0.032***	0.062***	0.086***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	1,340,806	1,348,256	1,345,121	1,363,079	1,341,684
Log Likelihood	-1,842,824.000	-1,641,516.000	-1,546,538.000	-1,440,409.000	-1,715,297.000
Akaike Inf. Crit.	3,685,656.000	3,283,041.000	3,093,084.000	2,880,825.000	3,430,602.000
Bayesian Inf. Crit.	3,685,704.000	3,283,089.000	3,093,132.000	2,880,874.000	3,430,650.000
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 2: Effect on student progress by domain in mathematics

	Dependent variable:				
	Difference in scores between early-second-grade and mid-first-grade				
	Addition	Arithmetic problems	Number line	Number writing	Substraction
	(1)	(2)	(3)	(4)	(5)
year 2020	−0.039***	−0.018***	−0.126***	−0.113***	−0.020***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.007***	0.001	0.065***	0.056***	−0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	1,345,025	1,379,336	1,361,735	1,345,057	1,335,555
Log Likelihood	-1,974,025.000	-1,983,149.000	-1,987,594.000	-1,841,827.000	-2,034,200.000
Akaike Inf. Crit.	3,948,058.000	3,966,306.000	3,975,197.000	3,683,662.000	4,068,409.000
Bayesian Inf. Crit.	3,948,106.000	3,966,354.000	3,975,245.000	3,683,711.000	4,068,457.000
Note:					*p<0.1; **p<0.05; ***p<0.01

The most affected areas in French concern writing skills: *Syllable writing* and *Word writing*. It can be seen in Figure 21 that the social factor play a crucial role during the lockdown for these domains: the correlation between academic performance and social context increases most for these two domains between cohort $\mathcal{C}_0$ and cohort $\mathcal{C}_1$. A similar phenomenon can be observed in mathematics with the domain *Number writing*: it is one of the most impacted of all and the one where the social gap increases the most during the lockdown. This shows that in normal times, school plays its role in learning to write: when it is lacking, students from the most disadvantaged social classes are mechanically the most affected and inequalities explode, resulting in a sharp drop in progression in this area compared to a standard year.

Learning to read, another keystone of the first grade curriculum, is also one of the areas in which the school succeeds in smoothing out some of social inequalities, since a similar phenomenon, but to a lesser extent can be observed with *Word reading* and *Text reading.*

Conversely, the domain least affected in French is *Spoken sentence comprehension.* Figure 21 shows that the correlation between social context and student performance is the highest for this domain, even in normal times. Moreover, its value does not change after the lockdown. This reflects the strong inequalities in language comprehension that preexisted to the pandemic crisis: students from high social backgrounds start first grade with a very high level of oral comprehension compared to the average, especially since oral comprehension leverages skills which include the development of vocabulary, that does not take place mainly at school, but rather at home, in the private sphere. The gap between pupils from advantaged and disadvantaged backgrounds is already so wide in normal circumstances that the lockdown has not made much difference: schools already struggle to compensate social inequalities in this domain. Similar conclusions can be drawn for the mathematical domain *Arithmetic problems*, the least affected by the pandemic crisis, and the most socially determined.

5 Conclusion

The health crisis caused by COVID-19, which began at the dawn of the year 2020, provoked a profound rupture in people's lives, the aftermath of which is still visible today. Very quickly, everyone had to adapt to new ways of organization, work, consumption and social life. While the economic short-term effects of this unprecedented crisis are widely studied, it is much more difficult to measure the long-term effects. What will be the remaining impact of this pandemic in 20 years? In 30 years? Will our society still bear the marks of this health crisis? To answer these questions, it is essential to look at the people who will make up a large part of the working population. These people are now children. In particular, the pupils who entered first grade in September 2019 experienced the full impact of this crisis in their first year of primary schooling, the year they learn to read, write and count. In France, the closure of schools between March and May 2020 deprived them of any social interaction outside of their homes for two months, and forced them to follow distance learning, according to a protocol that was not always efficient or functional, in conditions that were sometimes difficult (large families living in small areas, poorly equipped, without an internet connection, etc.). What will be the long-term consequences of this disrupted schooling for this generation of students? This is a broad question, and two years later it is not possible to answer it. However, we can begin to explore it by looking at the short-term effects.

This study is devoted to measure these effects, on a strictly academic level. Thanks to the data from the national assessments, it is possible to examine how school closures affected student progression between mid-first-grade and early-second-grade, in French and mathematics. Our analysis highlights the fact that academic progress, relative to a normal year, dropped for students who experienced the lockdown during first grade. The negative impact is greater in French, where it represents 22 % of the the standard deviation of the control population distribution, than in mathematics, with 11 % standard deviation. From a more general perspective, the literature on school interruptions usually shows that learning losses are greater in mathematics than in reading (Kuhfeld et al. [8]). However, recent works on school closure due to covid crisis also measure a greater learning loss in language than in mathematics (Maldonado et al. [7], Schult et al. [10]). Unlike school closures studied in the past, the lockdown was accompanied by the introduction of distance learning during the lockdown. One possible explanation for this phenomenon is that mathematics is more easy to teach at a distance than reading or writing, where direct teacher-student interaction is more fundamental.

In both disciplines, not all students were affected with the same intensity. There is evidence that it has deepened a number of pre-existing inequalities: across disciplines and in the majority of assessed domains, students who were already in the weakest positions suffered more from the lockdown than their peers who benefited from more favorable conditions. Students who started first grade with the greatest difficulties found themselves the most helpless in the face of school closures. In addition, this period without school was mostly detrimental to students in priority education, and somewhat less so to those in private education, compared to students enrolled in public education. However, at equivalent social context of the schools, pupils in priority education suffered slightly less than those in public education, which shows that the priority education policy is at least partially successful. Conversely, while private school students appear to be less affected by school closures, they are no better off than public school students at equivalent social levels, and even slightly worse for those who returned to school before the summer break. However, this last result should be seen in the light of the specific social profile of students who did not return to school between the reopening in May and the summer vacations. More precisely, the estimated effects on students who did not return to school before summer not only measure the impact of a longer period without school but also contain the effect of specific characteristics of this particular population.

Our study also revealed structural inequalities that schools fail to address, even under normal circumstances. This is the case, for example, with regard to language comprehension, an area that is crucial to the future social integration of students. Indeed, when analyzing the results by skills area, the French domain least affected by school closures concerns oral comprehension, which shows its profoundly social dimension: school education does not manage to compensate for the strong inequalities in language comprehension. On the contrary, we notice that domains related to writing are the most affected, whether in French (writing syllables or words) or in mathematics (writing numbers). Moreover, the progression in reading is also heavily impacted. This underlines the importance of school in these two fundamental learnings of first grade: without school, competences in this field fall drastically especially for disadvantaged social categories.

What can be done to reduce these inequalities that have been exacerbated by the crisis? Will these children catch up on what they missed that year? How long will it take them? Will the gaps that the crisis has widened be permanent for this generation? All these questions remain open and deserve to be addressed. This would require further investigations to track these students achievements' in higher grades, which might be possible thanks to the French national assessments taken by each student entering high school.

Bibliographie

References

[1] Davezies, L. Modèles à effets fixes, à effets aléatoires, modèles mixtes ou multi-niveaux : propriétés et mises en oeuvre des modélisations de l'hétérogénéité dans le cas de données groupées. *Série des documents de travail de la Direction des Études et Synthèses Économiques* (2011).

[2] D'Haultfoeuille, X. Econometrics 1: Lecture notes. *École Nationale de la Statistique et de l'Administration Économique*.

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

[4] Garbinti, B. Econometrics: Lecture notes. *École Nationale de la Statistique et de l'Administration Économique*.

[5] Givord, P., and Guillerm, M. Les modèles multiniveaux. *Série des documents de travail « Méthodologie Statistique »* (2016).

[6] James, G., Witten, D., Hastie, T., and Tibshirani, R. *An Introduction to Statistical Learning: with Applications in R.* Springer, 2013.

[7] Maldonado, J. E., and De Witte, K. The effect of school closures on standardised student test outcomes. *British Educational Research Journal 48*, 1 (2022), 49–94.

[8] Megan Kuhfeld, James Soland, B. T. A. J. E. R. J. L. Projecting the potential impacts of covid-19 school closures on academic achievement.

[9] Murat, F. Discussions méthodologiques sur les différentes méthodes économétriques pour estimer l'effet d'un traitement - annexe 8 du document de travail sur l'Évaluation de l'impact de la réduction de la taille des classes de cp et de ce1 en rep+ sur les résultats des élèves et les pratiques des enseignants. *Direction de l'évaluation, de la prospective et de la performance, Ministère de l'Education Nationale* (2020).

[10] Schult, J., Mahler, N., Fauth, B., and Lindner, M. A. Did students learn less during the covid-19 pandemic? reading and mathematics competencies before and after the first pandemic wave. *School Effectiveness and School Improvement 0*, 0 (2022), 1–20.

A Details about assessed areas and items

Discipline	Domain	Number of items		
		Common	2018	2019
French	Grapheme/phoneme correspondence	7	10	10
	Letter recognition	7	7	7
	Letter/non-letter discrimination	0	4	0
	Consonant string comparison	12	24	12
	Phoneme manipulation	11	15	15
	Syllable manipulation	15	15	15
	Spoken sentence comprehension	14	14	14
	Spoken text comprehension	11	18	11
	Spoken word comprehension	15	15	15
Mathematics	Number recognition	10	10	10
	Number writing	11	11	11
	Number line	6	6	6
	Number comparison	40	60	40
	Enumeration	8	8	8
	Assembly reproduction	0	0	8
	Arithmetic problems	3	6	6

Table 3: Domains in the early-first-grade assessments in September 2018 and 2019

Discipline	Domain	Number of items		
		Common	2019	2020
French	Grapheme/phoneme correspondence	7	10	10
	Phoneme manipulation	7	12	12
	Syllable writing	10	10	10
	Word writing	8	8	8
	Pseudoword reading	0	6	0
	Word reading	6	6	6
	Text reading	4	4	4
	Sentence reading comprehension	0	0	8
	Spoken sentence comprehension	14	14	14
Mathematics	Number writing	10	10	10
	Number comparison	40	40	40
	Number line	10	10	10
	Addition	7	7	10
	Substraction	6	7	10
	Arithmetic problems	5	5	5

Table 4: Domains in mid-first-grade assessments in January 2019 and 2020

Discipline	Domain	Number of items		
		Common	2019	2020
French	Syllable writing	12	12	12
	Word writing	12	12	12
	Word reading	61	73	61
	Text reading	2	4	2
	Read sentence comprehension	10	10	10
	Read text comprehension	8	8	8
	Spoken word comprehension	15	15	15
	Spoken sentence comprehension	15	15	15
Mathematics	Number recognition	10	10	10
	Number representation	16	16	16
	Number writing	10	10	10
	Number line	15	15	15
	Addition	4	7	8
	Substraction	5	8	7
	Mental calculation	10	10	10
	Assembly reproduction	8	8	8
	Arithmetic problems	6	6	6

Table 5: Domains in early-second-grade assessments in September 2019 and 2020

B Study on missing students ID

During the 2018-2019 academic year, a substantial number of schools did not fill out national student IDs of their first graders. This makes it impossible to connect the first-grade student ID to the second-grade student ID for 13 % student in this cohort, who are then removed from the scope of our study. Therefore, the first cohort consists of 669,000 students and the second of 769,000 (Figure 23). In this appendix, we examine whether the two cohorts are still comparable despite the missing students.

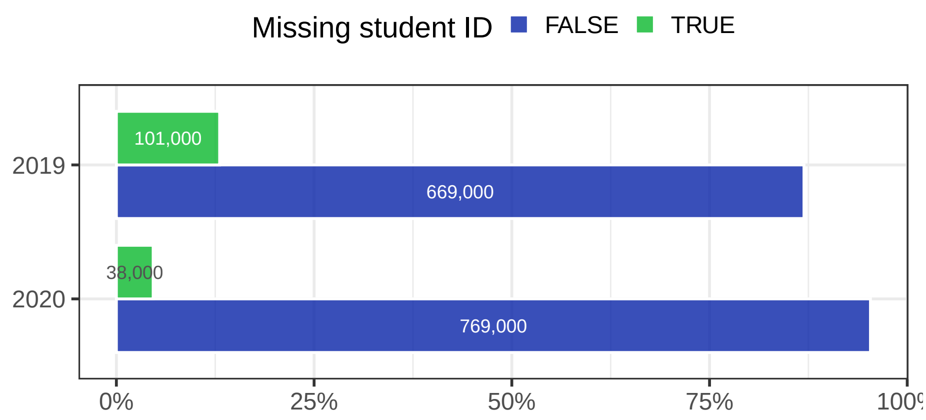


Figure 23: Number of students in the first cohort (2019) and second cohort (2020) whose ID was or was not found

Although the schools that did not correctly report the national student IDs are primarily in the private sector, the distribution of the remaining students in the first cohort across the three school sectors is comparable to that of the second cohort, as it can be seen in Figure 24.

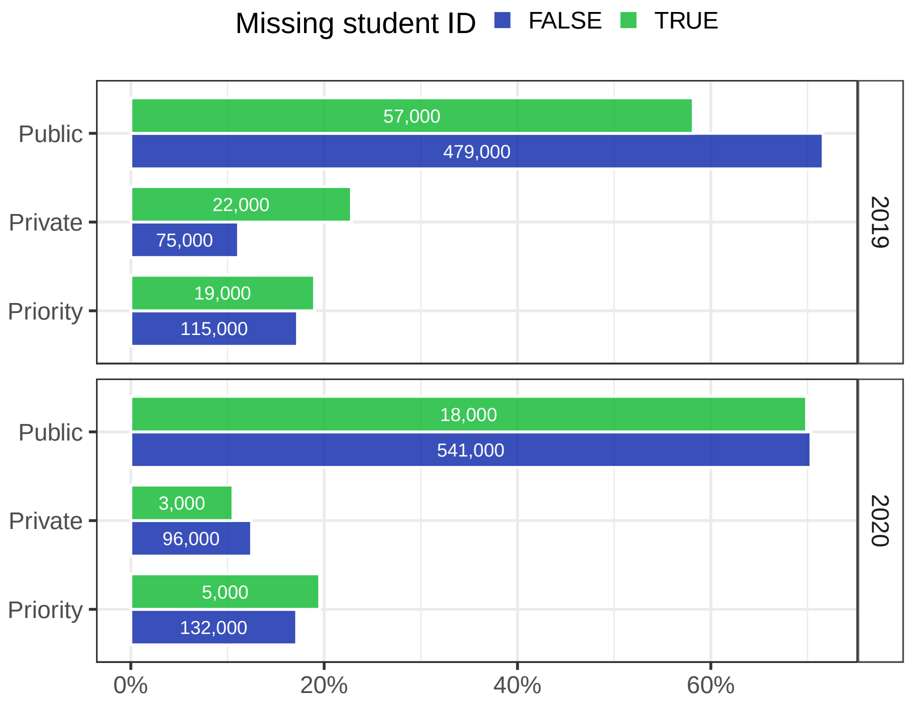


Figure 24: Proportion and number of students by school sector for both cohorts

Looking at the school social position index in Figure 25, its distribution among students is comparable between the two cohorts.

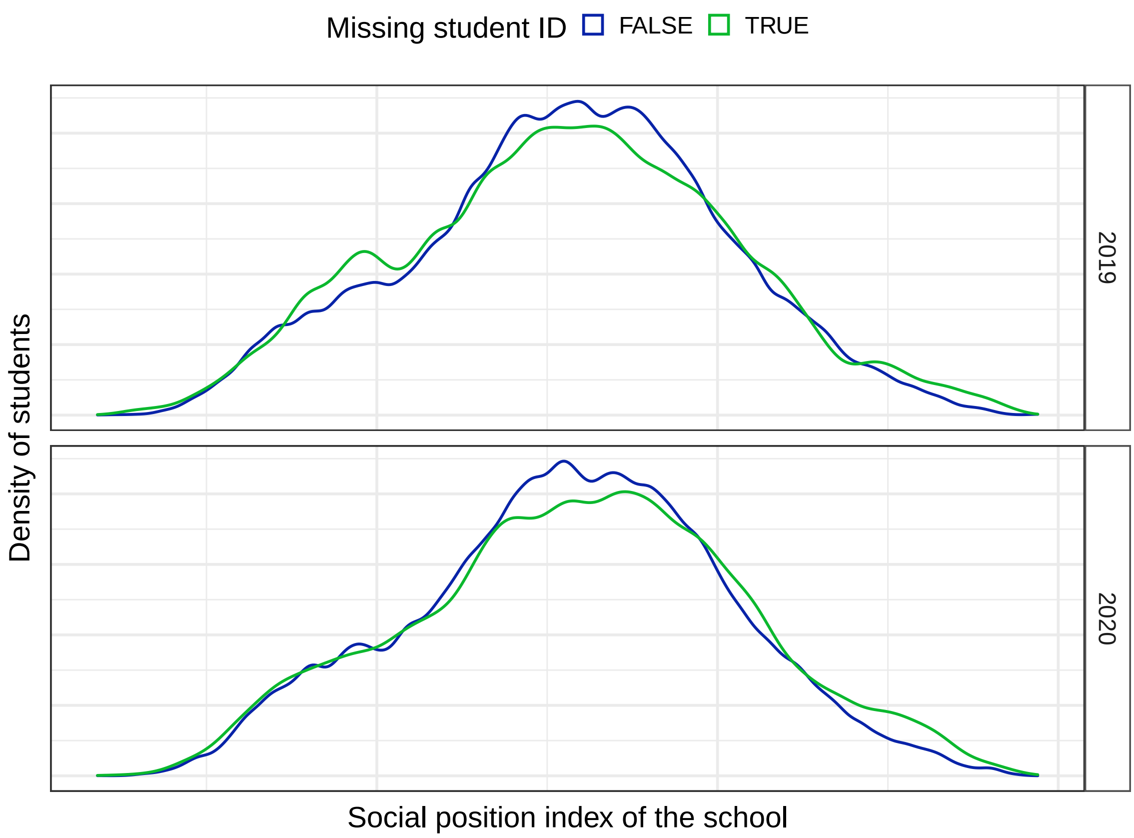


Figure 25: Distribution of social position index amongst the two cohorts of students

All of this information leads us to conclude that the two cohorts are still comparable despite the exclusion of students without academic ID.

C Visualization of the PCA results

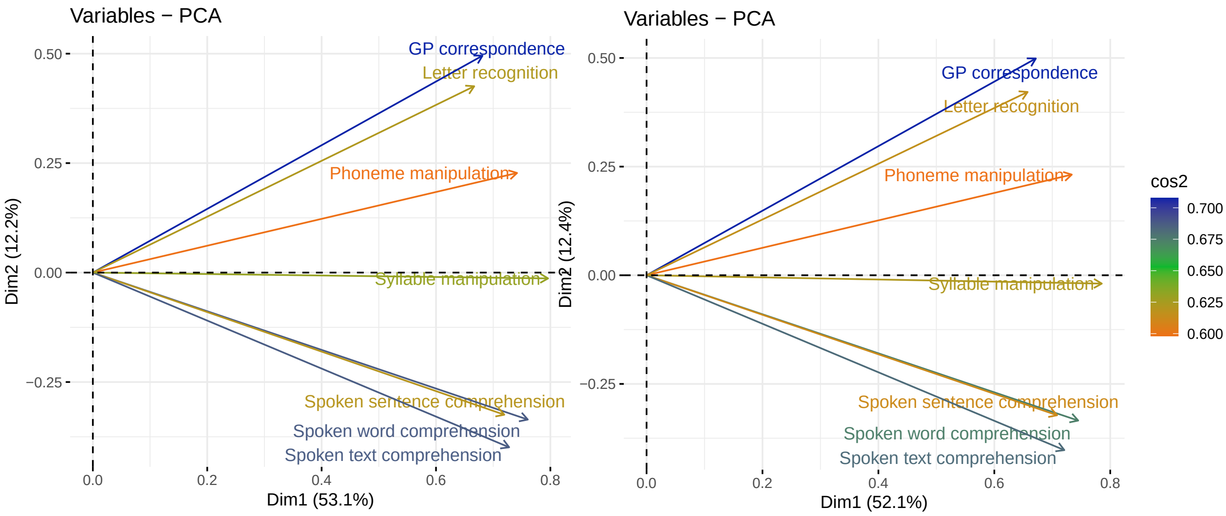
We perform a PCA to extract the first dimension that best describes students' abilities in the discipline. Our data contains a small proportion of students with at least one missing domain score, as shown in Table 6 below.

Assessment	French		Mathematics	
	Control	Treatment	Control	Treatment
early-first-grade	3.5	3.3	1.0	2.4
mid-first-grade	6.6	6.1	5.0	4.0
early-second-grade	2.6	4.1	2.2	3.2

Table 6: Percentages of students with at least one missing domain score

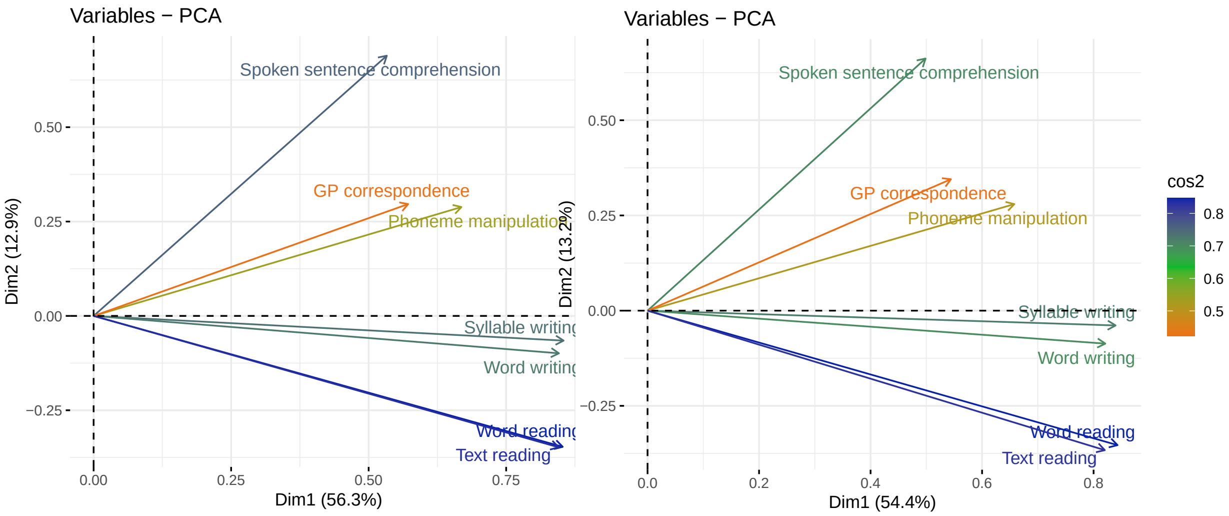
To deal with the missing domain scores, we use two methods: domain score imputation and student deletion. This appendix presents the results of both methods. Each of the following figures, from Figure 26 to Figure 31, shows the projections of the domain score variables onto the first two components calculated by PCA. On the left graph, the results are from the PCA run after imputing the missing scores and on the right, they are from the PCA when we remove the affected students from the data. The results are very similar between the two methods, so we choose the imputation method which has the advantage of not removing students for a few missing domains.

C.1 Projection of the French domains



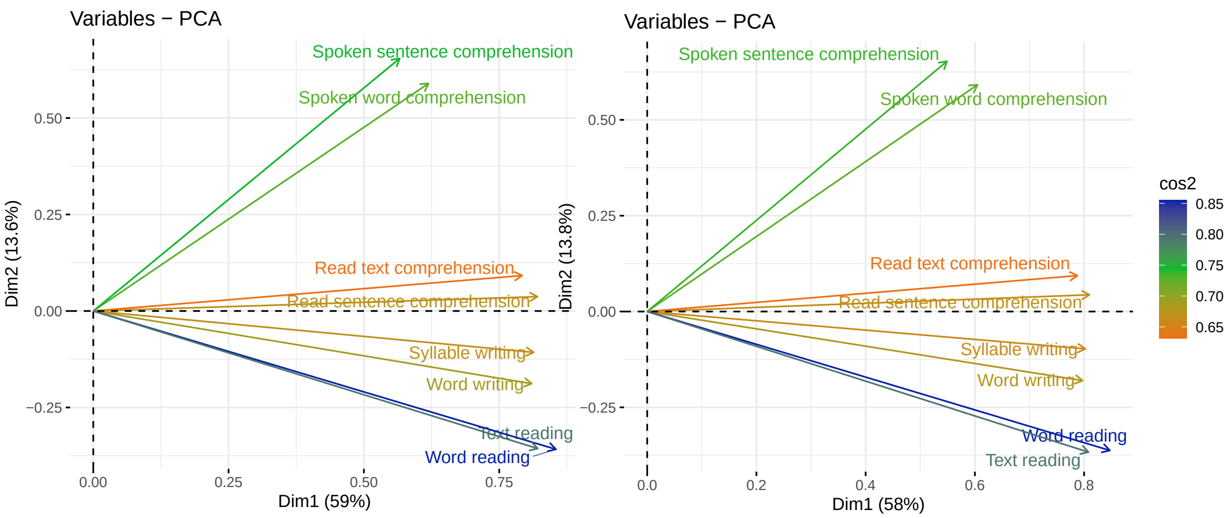
Imputation Deletion

Figure 26: Projection of domains for the early-first-grade French assessment



Imputation Deletion

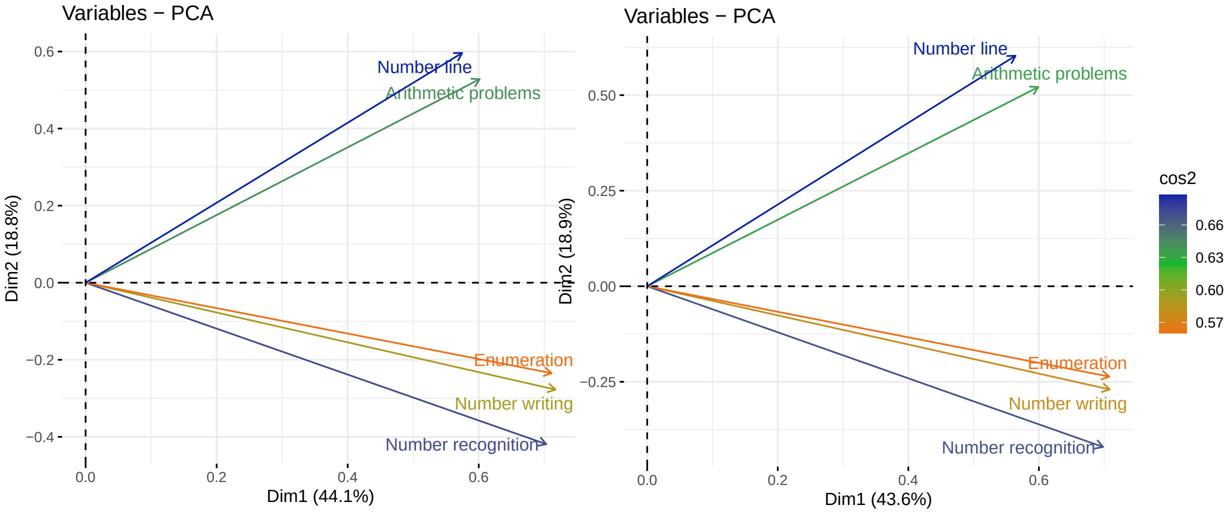
Figure 27: Projection of domains for the mid-first-grade French assessment



Imputation Deletion

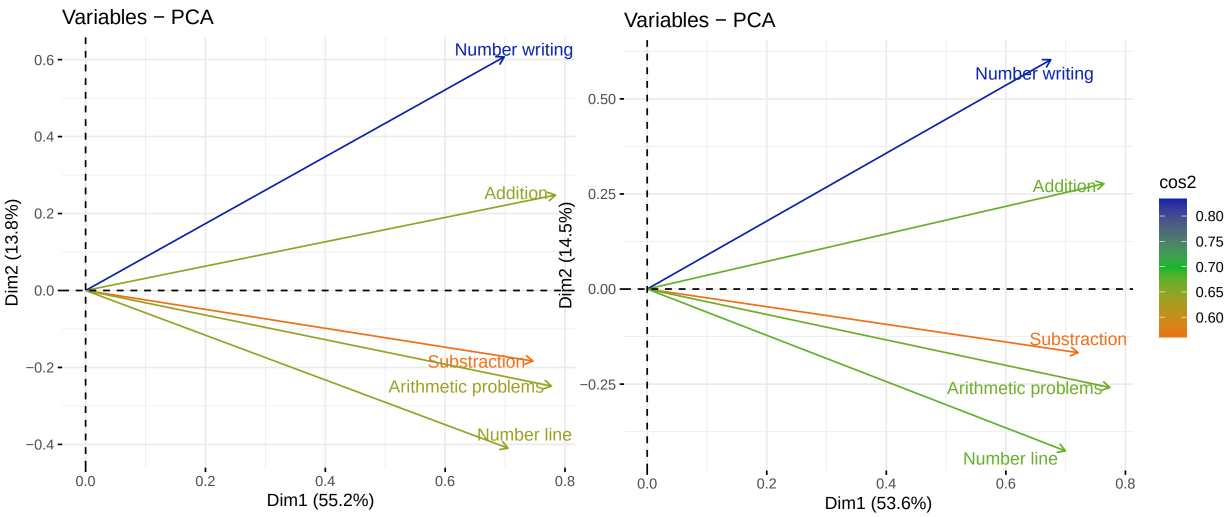
Figure 28: Projection of domains for the second-grade French assessment

C.2 Projection of the mathematical domains



Imputation Deletion

Figure 29: Projection of domains for the early-first-grade mathematical assessment



Imputation Deletion

Figure 30: Projection of domains for the mid-first-grade mathematical assessment

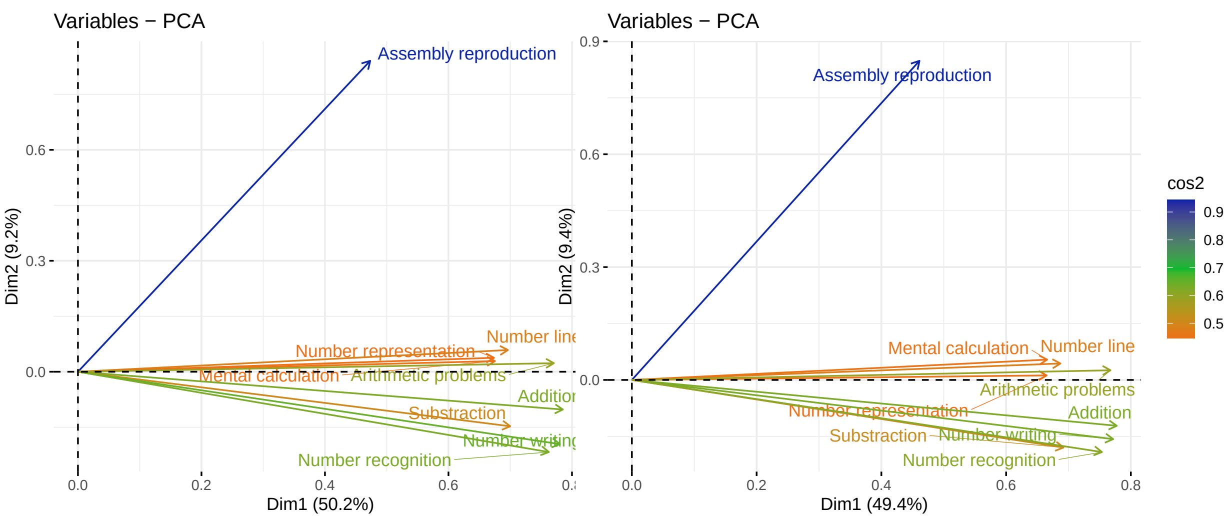


Figure 31: Projection of domains for the second-grade mathematical assessment

D Evolution of student performance by domain

Figure 32: Average student scores by domains in French

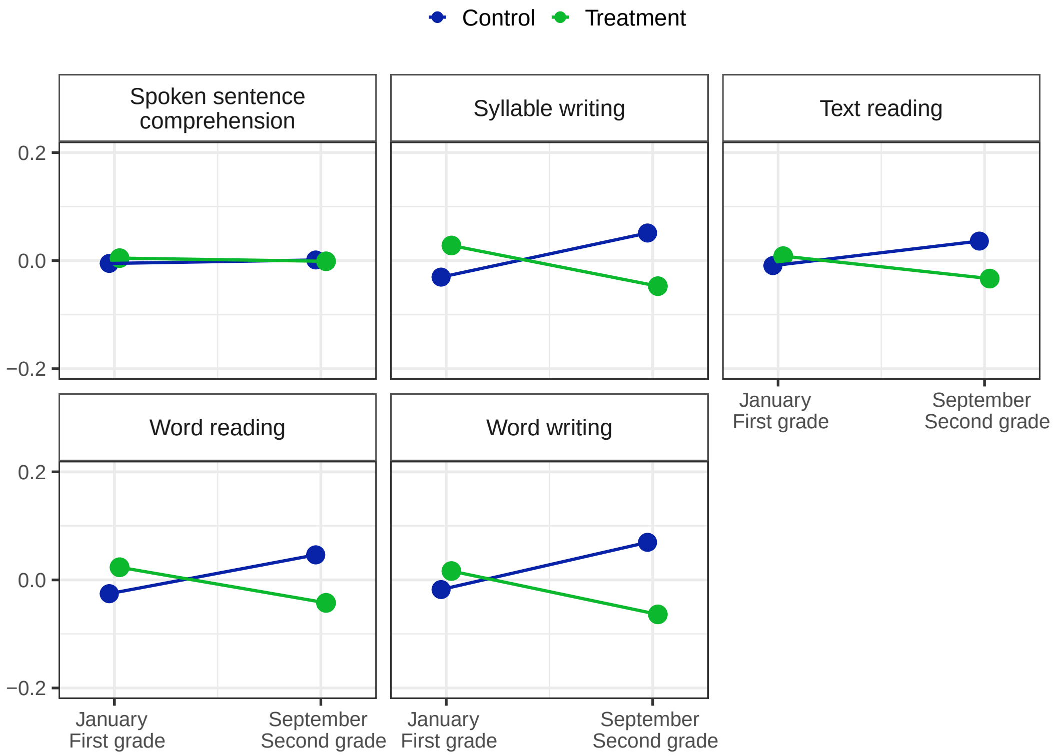
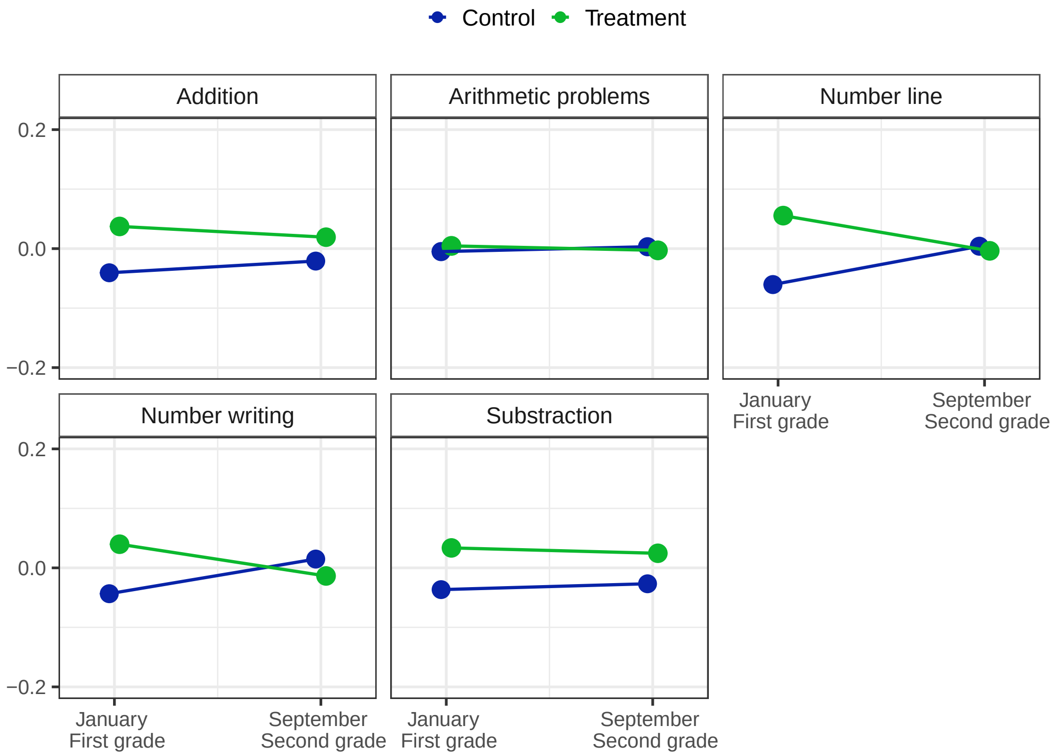


Figure 33: Average student scores by domains in Mathematics



E Effect of school closures estimated by OLS models

Table 7: OLS in French

	Dependent variable:			
	Difference in scores between early-second-grade and mid-first-grade			
	(1)	(2)	(3)	(4)
year 2020	−0.144***	−0.140***	−0.154***	−0.152***
	(0.001)	(0.002)	(0.002)	(0.002)
gender female		0.054***	0.054***	0.063***
		(0.002)	(0.002)	(0.002)
sector priority		−0.168***	−0.122***	−0.135***
		(0.002)	(0.003)	(0.003)
sector private		0.024***	0.005*	0.009***
		(0.002)	(0.003)	(0.003)
SPI			0.032***	0.043***
			(0.001)	(0.001)
early-first-grade				−0.046***
				(0.001)
year 2020:gender female		0.008***	0.008***	0.001
		(0.002)	(0.002)	(0.002)
year 2020:sector priority		−0.084***	0.009**	0.021***
		(0.003)	(0.004)	(0.004)
year 2020:sector private		0.044***	0.011***	0.009**
		(0.003)	(0.003)	(0.004)
year 2020:SPI			0.061***	0.052***
			(0.001)	(0.001)
year 2020:early-first-grade				0.041***
				(0.001)
Constant	0.077***	0.077***	0.074***	0.071***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	1,375,640	1,375,640	1,304,589	1,269,095
Adjusted R^2	0.013	0.033	0.041	0.043
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 8: OLS in mathematics

	Dependent variable:			
	Difference in scores between early-second-grade and mid-first-grade			
	(1)	(2)	(3)	(4)
year 2020	−0.089***	−0.081***	−0.091***	−0.088***
	(0.001)	(0.002)	(0.002)	(0.002)
gender female		−0.128***	−0.127***	−0.124***
		(0.002)	(0.002)	(0.002)
sector priority		−0.047***	−0.001	−0.007**
		(0.002)	(0.003)	(0.003)
sector private		0.049***	0.028***	0.030***
		(0.003)	(0.003)	(0.003)
SPI			0.032***	0.038***
			(0.001)	(0.001)
early-first-grade				−0.038***
				(0.001)
year 2020:gender female		0.009***	0.008***	0.008***
		(0.003)	(0.003)	(0.003)
year 2020:sector priority		−0.096***	−0.023***	−0.023***
		(0.003)	(0.004)	(0.004)
year 2020:sector private		0.025***	−0.004	−0.004
		(0.004)	(0.004)	(0.004)
year 2020:SPI			0.050***	0.047***
			(0.002)	(0.002)
year 2020:early-first-grade				0.014***
				(0.001)
Constant	0.048***	0.113***	0.110***	0.108***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	1,378,131	1,378,131	1,306,909	1,264,558
Adjusted R^2	0.004	0.014	0.019	0.021
Note:			*p<0.1; **p<0.05; ***p<0.01	