2021-11

Working paper. Economics

ISSN 2340-5031

EFFECT OF CLASS SIZE ON STUDENT ACHIEVEMENT IN THE COVID-19 `NEW NORMAL´.

Jesús M. Carro, Pedro Gallardo.

uc3m | Universidad Carlos III de Madrid

Effect of Class Size on Student Achievement in the COVID-19 'New Normal'*

Jesus M. Carro† Pedro Gallardo ‡

06 September 2021

Abstract

The COVID-19 pandemic implied measures, such as school closures at the outbreak of the pandemic, that negatively affected children's human capital. However, in some places, the situation later brought a reduction in class sizes, in order to avoid the spread of SARS-Cov-2 at schools. We exploit this unexpected event to evaluate the effect on school performance of a sizable class size reduction implemented in Spain, when schools were reopened. We find a positive and significant effect of the class size reduction of 0.11 standard deviations on overall students' performance. Given the situation and nature of our data, we interpret our estimates as a lower bound for the true effect of the reduction in class size. Our conclusion is that the reduction in class size served, on average, to at least compensate the other negative effects for learning. Our findings also point out at the importance of evaluating the quality of the new and unexperienced additional teachers that need to be hired when implementing a general reduction of class size.

JEL classification: I2; H4

Keywords: Class size, SARS-COV-2, COVID-19, student achievement, students/teacher ratio.

*We are grateful to the 'Consejeria de Educación y Empleo' of the regional Government of Extremadura (Spain) and the employees there that authorized and helped accessing the data. All remaining errors and responsibility for this work are our own. The first author gratefully acknowledge research funding from the Agencia Estatal de Investigación del Gobierno de España, grant RTI2018-095231-B-I00, and Comunidad de Madrid (Spain), grant EPUC3M11 (V PRICIT). There is no other finantial sources nor potential conflict of interest to declare.

†Corresponding Author. Department of Economics. Universidad Carlos III de Madrid. Calle Madrid, 126. Getafe, Spain. E-mail: jcarro@eco.uc3m.es

‡Universidad Carlos III de Madrid. E-mail: 100366843@alumnos.uc3m.es

1 Introduction

The spread of the SARS-COV-2 virus brought school closures to many countries, as a way to control the infection, especially in 2020. This, indeed, had consequences for human capital accumulation (Fuchs-Schündeln et. al., 2020), learning (Engzell et.al., 2020), and other outcomes (Takaku and Yokoyama, 2021). In many countries schools reopened under strict sanitary measures and surveillance. It has been shown that, under those conditions, they were not a source of an increase in COVID case numbers (Isphording et.al, 2021). Spain is a leading example for all those situations among European countries. Firstly, it has been severely affected by the pandemic from the beginning of the spread of the decease in Europe, resulting in one of the highest excess of mortality rate. In one year of pandemic, from March 2020 to February 2021, the excess in the number of deaths with respect to the average number of deaths in preceding years was 23% in Spain, in comparison to 21% in the UK, 20% in Italy, 14% in France and 4% in Denmark, to mention just a few examples of countries affected with different intensities.[1,2]

Secondly, given the strong initial intensity of the spread of the pandemic, in mid-March 2020 a strict and long confinement and suspension of all nonessential activities (that could not be done from home) was decreed by the Spanish government. In particular schools were closed from mid-March 2020 to the beginning of the next school year, in September 2020. That means that schools were closed for a third of the 2019/20 school year, without having had time to prepare schools, teachers, and much less students and their homes, for online learning.

Thirdly, in September 2020 all schools were opened and they remained opened for the entire school year, even when some other activities were again temporary restricted as new waves of the infection were hitting different Spanish regions. The opening of the schools was done under strict monitoring and sanitary conditions, which included face-masks, intensive hygiene, ventilation, and social distance. Among the public policy measures intended to keep interpersonal distances, the main one in terms of its cost was a sizeable reduction in class-sizes implemented at several education grades in many regions across the country. In this paper we study the academic effects of such unprecedented class-size reduction, in a context where other measures and aspects of the sanitary crisis difficulted learning. What was the effect of the reduction in class-size on

[1] The definition of those excess of mortality rates, or P-scores, is the percentage difference between the number of deaths in (March 2020 to February 2021) and the average number of deaths in the same period over the years 2015–2019. See Ritchie et. al. (2021) for more details on this measure.

[2] These percentages have been calculated using data from the Human Mortality Database (HMD) Short-term Mortality Fluctuations project, and the World Mortality Dataset (WMD), accessible online at https://www.mortality.org/, and at https://github.com/akarlinsky/world_mortality.

student academic results? Was it insignificant or was it positive enough as to compensate the learning losses due the school closures, such that students have overall better academic results? Is reducing class size a good public policy to compensate losses in academic performance from other determinants?

The effect of class size on students outcomes has been subject to controversy for several decades. See, for example, Dustman (2003), the editorial of the special Feature on this debate at *The Economic Journal*. Experimental or quasi-experimental data, and careful analysis of detailed observational data has shown positive and statistically significant impact of both a reduction on class size and an increase on per-pupil spending (of which the students/teacher ratio is the main determinant). A few examples of those are the following. Krueger (1999) uses the project STAR experimental data and finds a positive and significant effect of reducing class size on student performance. Jackson et.al. (2016) uses quasi-experimental variation generated by school finance reforms to study the effect of per-pupil spending and finds a positive and significant effect on several outcomes. Dustmann et. al. (2003) studies reasons for the contradictory results when using non-experimental data, which will allow them to have more efficient estimates. Using it with detailed observational data Dustmann et. al. (2003) finds positive and significant effects of reducing class-size on future wages. Finally, Krueger (2003) revisits previously done meta-analysis of the literature, like Hanushek (1997), to show a reason for the contradiction between previous conclusions from summarizing the literature and the experimental evidence. According to the summary of the literature and weights of the different estimates done by Krueger (2003), there is a systematic positive relation between school resources and student achievement.

Despite of those results, class size has remained controverted as an efficient public policy to improve students performance and education quality. Being the effects of class size small in quantity and its financial cost large, it is important to have more estimates, such that more cost-benefit calculations relevant to different places and situations could be done. At the same time, there are questions about its efficiency when compared with other measures aimed, for instance, to improve teacher quality, as Hanushek (2003) argue. Having more estimates of class-size effect will help resolve these issues.

While the COVID-19 pandemic implied an anticipated negative shock to human capital, it also triggered an exogenous reduction in class size at an unprecedented scale in a country that had tended to increase its students/teacher ratio during the preceding 15 years. This creates a unique setting to shed more light in the effect of class-size reduction as a way to increase education quality. Furthermore, our analysis raises the question about the quality in the hiring of new teachers needed to implement a class size reduction public policy at a large scale. Experimental implementations and evaluations of class size reductions can avoid the general equilibrium effect of increasing the aggregate demand for teachers because their scale is smaller. The extra teachers needed can come from the existing set of experienced teachers in the market. What we observe and evaluate in this paper is a general public policy that increases the number of groups for some grades all across the country. There were some re-organization within the system and schools, but eventually new and unexperienced teachers had to be hired via a less strict selection procedure. This may implied a separate negative effect if those new teachers are of worse quality. This is an aspect mostly overlooked in the literature so far, and needed of further research.

The rest of paper is organized as follows. Section 2 presents the institutional background and the data. Section 3 specifies the empirical model and assumptions. Section 4 presents the results, and Section 5 concludes.

2 Institutional Background and Data

As said, in Spain schools were closed from March 2020 until the beginning of the next school year in September 2020. For the school year 2020/21 sanitary and organizational instructions were given to prevent the spread of the infection in the classrooms. The specificities of these instructions and the implementation of them was at the hands of the regional governments, in our case the government of Extremadura. Of interest to our study is the instruction that in each class an interpersonal space of 1.5 meters between students in the room must be guaranteed (Consejería de Educación y Empleo, 2020). If ordinary classrooms did not meet such condition, schools would receive extra funding and personnel to raise the number of groups.[3] This would reduce the number of students per group and, hence, it would guarantee interpersonal space between its members even in a standard classroom.

Other sanitary measures such us mandatory face-masks, hands-hygene, and ventilation were in place. In-person classes were to be preserved as much as possible. If any member of the class tested positive for coronavirus, a reduced number of students that were sitting around to that

[3] In the entire region this resulted in the hiring of new teachers, increasing the total number of teachers working. See Junta de Extremadura (2020).

student were quarantined. The rest the group would continue to attend in-person classes. Individual students were also quarantined if any member of his/her house was tested positive, but this quarantine would not be extended to any other student in his/her classroom unless the student itself was tested positive.[4] Other aspects, such as the academic calendar, teaching hours, subject curriculum, and evaluation criteria were to remain the same as in any other year.[5]

The data was collected by the second author during the second half of the school year 2020/21, with the authorization and supervision of the 'Consejería de Educación y Empleo', the governmental department in charge of Education in the region of Extremadura. Given the restrictions for accessing into schools premises and extra workload for teachers and principals during this period, together with research time and resources constrains, only two schools were contacted and accepted to cooperate. Both are urban schools in a town from Extremadura, Spain. School I is a public Primary school and School II is a private but publicly funded school. These are the two (out of three) main types of schools and together school more than 97% of all primary school students in the region of Extremadura. The third type are non-subsidized private schools that serve no more than 3% of the student population at compulsory levels and operate under much less regulations.

The focus of our study were students in years 5 (Y-5) and 6 (Y-6) because they are in the final years of primary education. At these levels, grades are more comparable across years than the first three years of Primary Education, besides different rules regarding social distancing and other anti-covid measures were in place for the first years of Primary education. In year 2020/21 School I had five reduced groups in Y-5 and three in Y-6. School II had three groups per grade in Y-5 and Y-6. We use the term "reduced groups" to refer to the groups with smaller number of students that had to be made as a consequence of application of the social distancing requirement explained above. Table 1 presents the number of students, the average class-sizes per school and cohort, and how it changed from 2019/20 to 2020/21.As can be seen in the table, the reduction on class-sizes is between 33% (for most of the students) to 40% percent.

Important for our analysis, the redistribution of students among the new groups was made

[4] Overall, 33% of students were quarantine for at least a week during the first term of year 2020/21.

[5] Note that in the last term of 2019/20, special instructions were given by the education authorities regarding the grading of that particular term because schools were closed. For that reason the last term grades of the year 2019/20 are not comparable to grades in any other year. This was not the case in 2020/21 were the evaluation methodology and criteria were not modified and, in this respesct, are comparable to any other term and year but the last term of 2020/21.

Table 1: Number of groups and reduction in class size.

		Cohort of students defined by grade in 2020/21			
		School I		School II	
		Y-5	Y-6	Y-5	Y-6
Number of students		82	49	50	60
Number of groups	2019/20	3	2	2	2
	2020/21	5	3	3	3
Average class size	2019/20	27.3	24.5	25	30
	2020/21	16.4	16.3	16.3	20
	% Reduction	40%	33%	33%	33%

in a way that resembles random assignment. It is better explained with an specific case. School II had two Y-5 groups in school year 2019/20, Y5-1920A and Y5-1920B, each with 30 students. In order to allow for interpersonal space, in 2020/21 a new group is added such that those 60 students, now in Y-6, have to be reassigned into three groups, each having 20 students. They were ordered alphabetically within each existing group. The first 20 students in the alphabetical order of Y5-1920A were assigned to Y6-2021A; the first 20 students of Y5-1920B were assigned to Y6-2021B; and students in positions 20 to 30 in the alphabetical order in Y5-1920A and Y5-1920B formed a new 20-student class denoted Y6-2021C. This implies that, for each student, the peers in 20/21 are academically comparable to peers in 19/20, since the formation the new groups is not based on any determinant of the education production function.[6]

There are a total of 241 students in Y-5 and Y-6 in these two schools in year 2020/21. All of them were in that school in the previous academic year, 2019/2020. All of them but 10 students accepted to participate in the data collection. Among those that acepted to participate, two students are excluded because the information collected from them is incomplete. All these leave us with a sample of 229 students.

Data was collected from three different sources. First, students' grades for the first term (from September to December) of school years 2019/20 (before the pandemic) and 2020/21 were collected from the official academic records. From these we calculate our outcome variables: the Grade Point Average (GPA) of each student and Grades in Mathematics and Spanish. These grades, as the rest of grades in Spain, are a scale from 0 to 10. Second, students were surveyed about academic and family changes before (first term of 2019/20) and after (first term of 2020/21) the school closure for the pandemic. These include questions about study time

[6]Testing the mean peer's GPA in 2020/21 against the mean peer's GPA in 2019/20 confirms they are not statistically different for any group. These tests can be found in the Tables A.1 and A.2 in the Appendix.

at home, distractions in class, feeling worried, academic problems during the school closure, attending a private tutor, and labor status of the parents. Tables A.3 and A.4 in the Appendix contain some descriptive statistics of some of those variables. It should be noted from the descriptive statistics that, according to the student's answers, labor force status of the parents of these students has not changed as a consequence of the pandemic. Therefore this variable does not explain outcome changes and is not part of our explanatory variables.

As third source of information, the 13 teachers of the students in our data were surveyed about amount of individualized attention to students, classroom management, and their teaching and grading practices in 2020/21 (with the smaller class-size) with respect to 2019/20. None of them indicated that they had made changes in their evaluation methods, which confirms the situation described in the institutional background, and will allow us to compare grades of the first term across years. The other information from teachers will be used in our discussion about possible mechanisms that could produce the results we find. Teachers were also asked about their years of experience and whether they were new to the school in 2020/21. Only one teacher was new to the school and it was also the first year s/he was teaching. All the other teacher were highly experienced teachers. This again facilitates the isolation of the class-size effect from the effect of other determinants that may change with it. We will use this information later in our analysis by including it as a control variable.

3 Empirical Model and Identifying Assumptions

Consider the following education production function:

$$ACH_{i,t} = \alpha_i + \mathbf{I}'_{i,t}\boldsymbol{\beta}_I + \mathbf{F}'_{i,t}\boldsymbol{\beta}_F + \mathbf{S}_{i,t}\boldsymbol{\beta}_S + \beta_4 Classsize_{i,t} + \beta_5 Peers_{i,t} + \varepsilon_{i,t} \tag{1}$$

where i represent student, t represents time or, more precisely, $t = 1$ for the first term of the school year 2019/20 and $t = 2$ for the first term of the school year 2020/21; $ACH_{i,t}$ represents academic achievement, measured in our main specification by GPA, of student i at period t; $\mathbf{I}$ represent individual student inputs like number of hours of study or student's ability; $\mathbf{F}$ represents family factors like parent's education or income; $\mathbf{S}$ are school inputs including teacher's characteristics and quality; we separate class size, denoted $Classsize_{i,t}$, from the rest of school characteristics. In our data, $Classsize_{i,t}$ is going to take value one for the reduced groups, that is, $Classsize_{i,t} = 1$ for all i if $t = 2$, and 0 otherwise. $Peers_{i,t}$ represents average characteristics of student i's classmates that affect i's ACH; $\varepsilon_{i,t}$ are random shocks. Note that $\mathbf{I}$, $\mathbf{F}$, and $\mathbf{S}$ can include factors that we may observe as well as unobservables, and characteristics that can remain constant or vary over time. β_4 is our parameter of interest.

Assumption 1 : $\varepsilon_{i,t}$ is *iid* across i and t and independent of the other variables.

Assumption 1 is easily satisfied as long as it refers to random shocks, which is the case since in our notation unobservable characteristics are included in $\mathbf{I}$, $\mathbf{F}$, or $\mathbf{S}$. Of course, further assumptions are going to be needed regarding changes in these variables, and their validity and robustness of our results to them will be discussed. We will also have a separate discussion about whether the dummy $Classsize$ is capturing also the effect of other changes in inputs induced by the pandemic. However, for the reasons given, we maintain Assumption 1 through the rest of the paper.

Assumption 2 : Model (1) is applicable to all the periods in our sample.

Assumption 2 implies, not only that the education production function is stable over time, as usually assumed, but also that our $ACH_{i,t}$ are comparable over time, such that if both periods had the same inputs we would observe the same outcome. This requires the grading practices of teachers to be comparable in the two periods we consider. Both the regulations and the information reported by teachers are in line with this assumption for the two periods we observe. It also requires grades for the same student to be comparable across years, in particular Y-5 with Y-4 and Y-6 with Y-5.

Assumption 3 : (a) $Peers_{i,t} = Peers_{i,t-1}$ and (b) $\mathbf{F}_{i,t} = \mathbf{F}_{i,t-1}$

We make part (a) of this assumption because the students are the same in both years, and because of the random way the new groups of students are form in t with respect to $t-1$ that we have explained in Section 2. Part (b) comes from some variables in $\mathbf{F}$ being constant over time, like education of the parents, and from the students in our sample do not reporting any significant change in the labor status of their parents, as can be seen in Table A.4 in the Appendix.

First difference: same students comparison Our empirical specifications comes from taking a first difference in (1), and incorporating Assumption 3,

$$\Delta ACH_{i,t} = \Delta \mathbf{I}_{i,t}\boldsymbol{\beta}_I + \Delta \mathbf{S}_{i,t}\boldsymbol{\beta}_S + \beta_4 + \Delta\varepsilon_{i,t} \tag{2}$$

This is going to be our specification for the estimation part. Given that we have the same schools and students in $t-1$ and t, it could be further simplified to a constant,

$$\Delta ACH_{i,t} = \beta_4 + \Delta\varepsilon_{i,t} \tag{3}$$

if all the other individual and school inputs remain constant. In our data, we observe changes in $\mathbf{I}_{i,t}$ and $\mathbf{S}_{i,t}$, so (2) is our main equation. However, we estimate both (3) and (2), because the comparison between them can be informative about the potential effect of other non-observed variables, as we will explain.

4 Results

4.1 Estimates

Table 2: Estimates of equations (3) and (2).

	Dependent Variable					
	ΔGPA		ΔMath's Grade		ΔSpanish's Grade	
	(1)	(2)	(3)	(4)	(5)	(6)
β_4, Reduced Class Size	0.1316*** (0.0401)	0.1471* (0.0881)	-0.0742 (0.0778)	0.2654* (0.1552)	0.3188*** (0.0790)	0.3034* (0.1545)
Newly hired teacher		-0.2521 (0.1630)		-0.4317* (0.2395)		-1.2459*** (0.3006)
Study More		0.1373* (0.0698)		0.1950 (0.1288)		0.0771 (0.1286)
Quarantine, i.e. missed classes		-0.0820 (0.0862)		-0.3070* (0.1723)		0.3051* (0.1595)
Difficulties during confinement		-0.0529 (0.0874)		-0.3440** (0.1735)		0.1010 (0.1520)
Unease		-0.0164 (0.0876)		-0.2928* (0.1610)		0.0808 (0.1637)
Distracted more		0.0249 (0.0561)		0.1394 (0.1028)		0.0984 (0.1080)
N	229	229	229	229	229	229

Note: Standard errors are reported in parentheses. Asterisks indicate the estimate is significantly different from zero at *10%; **5%; ***1%. See main text in Section 4.1 for precise definitions of the covariates.

Table 2 contains the OLS estimates of equations (3) and (2). The covariates, $\Delta\mathbf{I}_{i,t}$ and $\Delta\mathbf{S}_{i,t}$, added are the following. An indicator variable '*Newly hired teacher*' that takes value one if the teacher of those students is the newly hired and unexperienced teacher we have in our sample. For the rest it takes value 0 since they all were present at the school the previous year and have many years of experience.[7] A variable 'Study More' that takes value 1 if i declared to study a higher number of hours in t than in $t-1$, 0 if it studied the same, and -1 if it studied less. 'Difficulties during confinement' takes value 1 if student i states that the distance learning in place during the confinement in Spring 2020 has affected negatively the student's preparation for the 2020/21 school year. Variable 'Unease' indicates if the student worries more about his or his family's health and wellbeing. Finally, 'Distracted more' takes value 1, 0, or -1 if the student indicates she gets more, equally, or less distracted when trying to pay attention in class in t than in $t-1$.

4.1.1 Effect over Grade Point Average

In all estimates we find and overall positive and significant effect of the reduction of class size, around 0.15 score points, which represents 0.11 standard deviations of the variable $GPA_{i,t}$. This is of smaller magnitude than effects found in project STAR –see for example Folder and Breda (1989)– but project STAR was for Kindergarten to third grade and our study is for fifth and sixth grades. Some of the explanatory variables, especially 'Distracted more' and 'Unease', are positively correlated, explaining the higher standard errors in the estimates when adding all of them together. The signs of the estimated effects of all the added explanatory variables are the expected ones, but of 'Distracted more'. However, the estimated effect of 'Distracted more' is small and not statistically significant. Actually, for the change in the GPA only 'Study More' has an effect that is statistically significant.

4.1.2 Separate effects for Maths and Spanish Language subjects

For Math and Spanish subjects the effect of reduction class size is larger in magnitude, representing 0.15 standard deviations for Math and 0.17 for Spanish. For Math several of the explanatory variables are statistically significant with the expected sign. In particular, difficulties during confinement with distance learning and 'Unease' have a large and significant negative effect for Math but not for Spanish nor the general GPA. Math is the subject most affected by the negative factor generated by the COVID pandemic. In contrast with that, learning of Spanish seems not to have suffered from learning from home (when a student is quarantine or during the confinement), nor from the other negative aspects generated by the pandemic.

[7] Rivkin et.al. (2005), for example, show that one additional year of experience does not affect teaching quality after the very first few years of experience.

4.1.3 Effect of new and unexperienced teacher

This is the larger effect that we find in Table 2. It is statistically significant for Math and Spanish, with the effect in the latter subject being 0.69 standard deviations. Even though the average effect for all students is positive, the estimated effect of the new teacher is greater in magnitude than the effect of the class size reduction for the group of students affected. The average GPA in 2019/20 of the students assigned to the group with the new teacher in 2020/21, which is 6.93, is a bit larger but statistically the same as the GPA in 2019/20 of the students assigned to the experienced teachers in the same school, which is 6.83. Therefore this effect is not due to the group assigned to the new teacher having worse previous performance than the other groups.

This is a very relevant variable for our study because it was induced precisely by the reduction on class size policy. If you increase the number of groups by about a third in an entire region, new teachers are going to be hired. Ignoring this consequence will lead to underestimate β_4, because the reduced class size dummy will take the negative effect of unexperienced new teachers. However, when evaluating the overall effect of the policy, not only the benefits of reducing class size have to be taken into account, but also the effect of the new hirings required to implement the policy at such a general scale.

There are two different reasons for the high negative effect of the new teacher. The first one is that this teacher has no experience at all and, as found in other papers, the first four or five years of experience have a positive impact on teaching quality, particularly the first year. The second reason may be that the new teachers that had to be hired at such a scale may be lower quality teachers. Under the first reason the effect is temporal and this event would have served to increase the human capital and training of new teachers that may be available for future years. Under the second reason, if the lower quality is a permanent characteristic, then the loss due to new teachers will have to be accounted for when evaluating the overall benefit of reducing class size as a general public policy aimed at increasing quality of education and human capital.

In our small sample of groups only one new teacher needed to be hired and it was an unexperienced teacher, so we cannot disentangle both effects. Furthermore, even with a bigger sample, new teachers would have being observed only for one year. We would need to observe them for several years to determine how permanent the negative effect is for them.

4.2 Assessing other potential confounding factors

4.2.1 Private tutor

A reaction from the side of the families to the negative effects of the school closure could be to hire private tutors for the 2020/21 school year to help the students outside school. If that were the case we would be incorrectly attributing to the class-size reduction what is due to more private tutors being hired at the same time class-sizes are reduced. We have information in our data about private tutors for each individual student in both pre and post pandemic periods. In 2019/20 18.8% of the students had private tutors. In 2020/21 6.14% ended having a private tutor and 8.74% started having one, leaving 21.3% of students with a private tutor in 2020/21. Taking into consideration all the changes, we cannot reject that the proportion is statistically the same in both years. Yet, since there is a slight increase, we have estimated model (2) adding this covariate too and obtain the same effect of reduced class size as in table 2. We do not report them because the estimated effect of private tutor is biased. The decision of having a private tutor in general, and of ending or starting having one is endogenous in the education production function. The relevant conclusion for our aim from those estimates is that our results on the effect of reduced class size are not driven by more students having private tutors.

4.2.2 Omitted variables

Any omitted variable that has remained constant from $t-1$ to t does not pose any problem to our estimates. However, there might be other changes for which we have not accounted, as this is not a truly randomized control experiment. In particular, the pandemic may have affected students performance in other ways than those we have been able to measure. Looking at the changes in our estimated effect when accounting for the observables we have, any other change goes in the direction of decreasing GPA and, therefore, incorrectly underestimating the benefits of Reducing Class Size if they are omitted. In fact, including covariates in our estimates increases and not decreases the estimated effect of class size. This is especially relevant for Math. If the bias when omitting our observables provides a guide for the bias from the potential unobservables, then, if anything, our estimate would be a lower bound and the benefits of reducing class size would be even higher.[8] Being a lower bound, in addition to the different grades mentioned above, might

[8] This argument follows the same idea as the selection on unobservable vs. selection on observables developed by Altonji et. al. (2005).

explain the smaller magnitude estimated compared with project STAR. Also, we are looking at outcomes after only a term with the reduced groups, and not at the end of the shool year where the accumulated effect could be higher.

4.3 Possible mechanisms

Given the positive and significant effect, we finish our study with some comments about the potential mechanisms that could produce that effect. This could guide future research. We use teachers's survey for this purpose. Teachers were asked seven questions. Four questions were about their interaction with students and behavior of them in class compared with previous years: (i) teachers get to know students' needs; (ii) provide individualized attention to students; (iii) students participation in class; (iv) teacher having to stop their lectures due to interruptions. Another question to teachers was whether they felt more motivated this than in previous years. The other two questions were about their workload: (vi) more/less/equal time spent preparing classes; (vii) amount of administrative duties.

Table 3: Number of teachers that reported doing some activities more/same/less than in the past.

	More	Same	Less
Know students needs	10	1	1
Individualized attention	8	3	1
Students participate in class	8	1	3
Lecture interruptions	4	3	5
Teachers motivated	6	2	4
Time preparing classes	7	5	0
Administrative duties	8	3	1

Table 3 summarizes the answers of the 12 teachers to those questions.[9] These answers suggest as a possible mechanism for the effect of reducing class size that teachers get to know better the students' needs, are able to provided more individualized attention to students and the students participate more in class. The last three questions point more towards the general feeling and workload of teachers than to a possible direct mechanism, but they can be affected by class size reduction and have an indirect effect. However they can also be affected by the other covid-related regulation and sanitary measures. The information collected points toward teachers not having less workload but yet being a bit more motivated than in previous years.

[9] Note that the new teacher could not answer those questions. It was his/her first time teaching, and hence could not compare his/her impressions to any preceding year.

5 Concluding remarks

We have found a positive and significant effect of class size reduction. The magnitude is not very large. Notwithstanding that, given that our data does not come from a randomized experiment, the unexpected shock that produced the reduction, and the nature and context (the COVID-19 pandemic) of our data, if anything our estimate is a lower bound of the effect, as we have discussed in section 4.2.2. Based on this, we can conclude that the reduction in class size has served to, at least, compensate other negative effects of the COVID in school performance.

Furthermore, our analysis is in a situation where the reduction has been implemented in the entire system and not only at a small scale, like in randomized control trials. A main difference with a small scale experiment is that new and unexperienced teachers had to be hired and this affect the final overall outcome. An interesting future research avenue, with a larger dataset of teachers, is to evaluate the impact of this on the distribution of teacher's quality and its evolution over time, separating permanent and transitory factors that determine teacher's quality.

References

[1] Altonji. J. G., T. E. Elder, and C. R. Taber (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools", *Journal of Political Economy*, Vol. 113, No. 1, pp. 151-184

[2] Consejería de Educación y Empelo (2020): "Guía General Para La Organización Y Desarrollo De La Actividad Educativa Para El Curso 2020/21 En Todos Los Centros Sostenidos Con Fondos Públicos De La Comunidad Autónoma De Extremadura" 'https://www.educarex.es/pub/cont/com/0064/documentos/guia_general_organizacion_desarrollo_actividad_educativa_2020_21.pdf'

[3] Dustmann, C. (2003): "The Class size debate and educational mechanisms: Editorial", *The Economic Journal*, Vol. 113, Iss. 485, pp. F1–F2

[4] Dustmann, C., N. Rajah, and A. van Soest (2003): "Class Size, Education, and Wages", *The Economic Journal*, Vol. 113, Iss. 485, pp. F99–F120

[5] Engzell, P., Frey, A., Verhagen, M.D., (2021): "Learning loss due to school closures during the COVID-19 pandemic". *PNAS* April 27, 2021 118 (17) e2022376118; https://doi.org/10.1073/pnas.2022376118.

[6] Folger, J., and C. Breda (1989): "Evidence from Project STAR about Class Size and Student Achievement", *Peabody Journal of Education*, Vol. 67 (1), pp. 17-33.

[7] Fuchs-Schündeln, N., Krueger, D., Ludwig, A., and Popova, I. (2020). "The Long-Term Distributional and Welfare Effects of Covid-19 School Closures". NBER Working Paper Series No. 27773, National Bureau of Economic Research.

[8] Isphording, Lipfert, and Pestel, (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 (2021) 104426.

[9] Jackson, C. K., R. C. Johnson, and C. Persico (2016): "The effects of school spending on educational and economic outcomes: evidence from school finance reforms", *The Quaterly Journal of Economics* (2016), Vol. 131, No. 1, pp. 157-218.

[10] Junta de Extremadura (2020): "Educación contratará 130 docentes más, con los que la ampliación total de la plantilla para hacer frente a la COVID-19 será de 744 plazas docentes." 'http://www.juntaex.es/comunicacion/noticia?idPub=31117#.YJ7aL6gzbIV'. [Online Resource]

[11] Krueger, A. (1999): "Experimental Estimates of Education Production Functions", *Quarterly Journal of Economics*, Vol. 114, No. 2, pp. 497-532.

[12] Krueger, A. (2003): "Economic Considerations and Class Size", *The Economic Journal*, Vol. 113, Iss. 485, pp. F34–F63.

[13] OECD (2021), The State of School Education: One Year into the COVID Pandemic, OECD Publishing, Paris, https://doi.org/10.1787/201dde84-en.

[14] Ritchie, H., Esteban Ortiz-Ospina, Diana Beltekian, Edouard Mathieu, Joe Hasell, Bobbie Macdonald, Charlie Giattino, Cameron Appel, Lucas Rodés-Guirao and Max Roser (2021) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/coronavirus' [Online Resource]

[15] Rivkin, S. G., E. A. Hanushek, J. F. Kain (2005): "Teachers, Schools, and Academic Achievement", *Econometrica*, Vol. 73, No. 2, pp. 417-58.

[16] Takaku, R., and I. Yokoyama, (2021). "What the COVID-19 school closure left in its wake: Evidence from a regression discontinuity analysis in Japan". *Journal Public Economics* 195, 104364.

A Appendix

Table A.1: Testing the mean peer's GPA in 2020/21 against the mean peer's GPA in 2019/20, School I

GROUP 20/21	GPA (VAR)	GROUP 19/20	GPA (VAR)	Difference	p-value, equality test
Y5A	8,3015 (1,6622)	Y4A	7,7989 (2,0650)	0,5026	0,2535
Y5B	7,6667 (1,6845)	Y4A	7,7989 (2,0650)	-0,1322	0,7705
Y5B	7,6667 (1,6845)	Y4C	8,0208 (1,4805)	-0,3542	0,4030
Y5C	7,9063 (1,9031	Y4C	8,0208 (1,4810)	-0,1146	0,7894
Y5D	7,3393 (2,3496)	Y4B	7,6741 (2,1161)	-0,3348	0,5036
Y5E	7,7768 (1,8394)	Y4B	7,6741 (2,1160)	0,1027	0,8231
Y6A	7,5547 (1,1895)	Y5A	7,3424 (1,7674)	0,2123	0,5884
Y6B	7,4375 (1,5907)	Y5B	7,337 (1,9440)	0,1005	0,8228
Y6C	7,9417 (1,7363)	Y5A	7,3424 (1,7674)	0,5993	0,1822
Y6C	7,9417 (1,7363)	Y6B	7,337 (1,9440)	0,6047	0,1864

Note: Variances are reported in parentheses.

Table A.2: Testing the mean peer's GPA in 2020/21 against the mean peer's GPA in 2019/2I, School II

GROUP 20/21	GPA (VAR)	GROUP 19/20	GPA (VAR)	Difference	p-value, equality test
Y5A	7,7647 (1,9334)	Y4A	7,6400 (1,0499)	0,1247	0,7544
Y5B	8,2059 (1,0233)	Y4B	8,0550 (0,7084)	0,1509	0,6158
Y5C	7,8516 (0,9296)	Y4A	7,6400 (1,0499)	0,2116	0,5082
Y5C	7,8516 (0,9296)	Y4B	8,0550 (0,7084)	-0,2034	0,4945
Y6A	6,9313 (1,3495)	Y5A	6,7542 (1,3917)	0,1771	0,6026
Y6B	7,1188 (2,2228)	Y6B	6,9792 (1,7469)	0,1396	0,7364
Y6C	7,1250 (1,9309)	Y6A	6,7542 (1,3916)	0,3708	0,3332
Y6C	7,1250 (1,9309)	Y6B	6,9792 (1,7469)	0,1458	0,7129

Note: Variances are reported in parentheses.

Table A.3: Descriptive statistics

	Mean	Stand. Dev.
GPA 2019/20	7.485	1.316
GPA 2020/21	7.617	1.361
Math's grade 2019/20	7.467	1.748
Math's grade 2020/21	7.393	1.753
Spanish's grade 2019/20	7.428	1.649
Spanish's grade 2020/21	7.747	1.801
Quarantine at home in 2020/21	0.3406	0.4750
Difficulties during confinement	0.4192	0.4945
Unease	0.3712	0.4842
With respect to 1st term of 2019/20, in 2020/21:		
Study More	0.5895	0.4930
Study Less	0.0742	0.2627
Study Same	0.3362	0.4735
Distracted more	0.1878	0.3914
Distracted less	0.5109	0.5010
Distracted same	0.3013	0.4598
Started with a Tutor	0.0877	0.2835
Ended with a Tutor	0.0617	0.2411
Still having a Tutor	0.1278	0.3346
Never have a Tutor	0.7181	0.4509

Table A.4: Reported labor status of the parents and changes before and after the pandemic.

	Proportion		
	2019/20	2020/21	Diference (standard error)
Mother works	0.7836	0.7706	-0.0130 (0.0234)
Mother not working	0.1299	0.1429	0.0130 (0.0179)
Missing response	0.0866	0.0866	0.0000 (0.0174)
Father works	0.8788	0.8788	0.0000 (0.0151)
Father not working	0.0476	0.0390	-0.0087 (0.0107)
Missing responde	0.0736	0.0823	0.0087 (0.0107)