DOES CLASS SIZE INFLUENCE STUDENT ACHIEVEMENT?

Summary
This policy brief presents the results of the most methodologically robust studies that estimate the impact of a reduction in class size on the student achievement and longer-run outcomes. Contrary to an idea that persisted for a long time, the effects are high, if they are compared with those of other educational policies conducted on a large scale and assessed rigorously. The absolute cost of such a policy justifies targeting it on the most disadvantaged pupils, but the investment is profitable whenever the salary gain for two years spent in a half-size class is greater than 1%. However many unknown quantities remain, in particular about how reduction in class size links up with teaching practices.

- A certain number of international research studies, most of which are recent, estimate the effects of a reduction in class size on the progress of the pupils by convincingly neutralising selection bias.
- Those studies show that halving the size of a class of 24 pupils by having two classes instead of one improves the mean performance levels of the pupils significantly, and it is even possible to see the long-term effects on academic pathways and occupational integration.
- These effects are observed even when the teachers are not given specific assistance or support with their teaching.
- The high cost of such a policy justifies targeting it on the most socially disadvantaged children, thereby working towards the goal of reducing inequalities.
- Cost-benefit calculations indicate that such a policy is profitable whenever it enables the future salaries of its beneficiaries to increase by 1%, through its benefits on academic pathways.
The impact of class size on pupils’ scholastic performance has been the subject of a considerable number of scientific articles and reports. Interest in this subject is fuelled both by the popularity of such policies with various stakeholders (teachers, and parents), and by the relative comparability of the effects measured in very different educational contexts. As the French Government starts to significantly reduce class sizes in the most disadvantaged schools, the results show us that although this policy is costly, it can reduce scholastic performance gaps when it is implemented in a targeted and intensive manner. The work and reports on France’s “priority education” policy generally emphasise that, since it was created, that policy has always failed to concentrate more resources per pupil in priority education establishments than in the rest of the education system (Bénabou et al., 2005; Cour des Comptes, 2016). By targeting the networks of schools classified as REP (Priority Education Network) and as REP+ (Priority Education Network Plus), the reductions in class sizes are making it possible, doubtless for the first time, to contribute much more to the most disadvantaged pupils.

Measuring the effects of class size: an empirical challenge

What is the effect of a reduction in class size on pupil performance? Paradoxically, it is difficult to answer this simple question. Imagine we are to observe the performance of school pupils in classes of various sizes. Merely by comparing the performance of pupils in the “large” and the “small” classes, we would often observe that the results of the pupils are better in the large classes.

Box 1: Methods used for measuring the effect of class size on pupil performance

It is not sufficient to compare pupils in “small” and “large” classes in order to measure the impact of class size on scholastic performance. Such a comparison is distorted by multiple selection biases. Below, we briefly describe the methods that make it possible to reduce such biases, or indeed to correct them totally.

“Value added” method. This method consists in comparing progress in the performance of the pupils in small and large classes, on the basis of standardised tests given at the start and end of the year. The main limitation with this approach is that it does not make it possible to correct the differences in progress that are not related to the dimensions of the initial level that are measured by the tests, but rather to other differences not observable between the classes of different sizes (such as the characteristics of the teachers or the characteristics of the pupils that are not picked up by the tests).

Randomised experiment. In order to neutralise the selection biases that distort comparison between small and large classes, the most convincing method consists in allocating the pupils and the teachers randomly to classes of different sizes. Comparing the results of the pupils assigned to the classes constituted in this way makes it possible to isolate the “pure” effect of class size on scholastic performance. To our knowledge, the only example of application of this method is the STAR experiment that was conducted in Tennessee in 1986.

Quasi-experiments. When a random assessment cannot be conducted, it is sometimes possible to use events or rules that naturally produce random variations, beyond the control of the researcher. For example, this might be the variation in class size that is induced by the size of the school cohort each year (Hooy, 2000), or it might be teacher recruitment forecasting errors that cause classes to be of larger or smaller size (Bressoux et al., 2009). We also include in this category the French non-randomised experiment in its most rigorous analysis (Bressoux and Lima, 2011).

Class-opening thresholds. This method can be considered to be a quasi-experiment or a “life-size” experiment insofar as it uses an administrative rule that is encountered in many countries to determine whether a new class should be opened in a school. Most education systems have a maximum number of pupils per class (that number is about 30 in France). When that number is exceeded, the establishment requires that a new class be created, inducing an “exogenous” reduction in the average size of the classes of any given grade or year depending on the total number of children enrolled in that grade or year.

Graph 1, opposite, taken from the study by Piketty and Valdeneire (2006), illustrates this phenomenon for the classes of CE1 (Year 3 or 2nd Grade) during the year 1998-1999. By comparing the pupils at schools whose class sizes are just above and just below the thresholds at which new classes must be opened, this quasi-random variation makes it possible to measure the causal effect of class size on scholastic performance.

Can we conclude from this that the reduction in class size affects performance negatively? Of course not. This misleading result is due to what researchers call “selection bias”: head teachers tend to place the least able pupils in the smallest classes whenever possible. Mere comparison of different class sizes thus teaches us absolutely nothing about the causal impact of reducing the number of pupils per class.

Testing at the start and end of the school year makes it possible to reduce this bias to some extent. It is then possible, instead of comparing the achievement or ability levels, to compare the progress observed in the large and small classes over the year (using the “value added” method, see box 1). This approach makes it possible to remove only some of the selection bias because the start-of-year tests will only measure a small part of the dimensions that determine the progress of the pupils, and that might have also influenced how they are distributed across the various classes. The vast majority of the studies on the effects of class size are based on this uncertain approach. Most of the studies taken into account in the classic summaries (meta-analyses) by Glass and Smith (1979, 1980) in the 1980s are of that nature. This also applies to the albeit very restrictive (eight studies) meta-analysis by Slavin (1989), and to the meta-analyses by Hanushek (1997) and by Hattie (2005). Such exercises are sensitive to the choice of the studies selected and to the weight given to their methodological soundness (Meuret, 2001; Krueger, 2003).
This policy brief is based on studies, most of which are recent, that have adopted strategies making it possible to neutralise this selection bias much more convincingly. They constitute the most reliable information currently available to researchers for estimating the impact of a reduction in class size (see Table 1). The most famous study is the randomised experiment “STAR” (which stands for “Student-Teacher Achievement Ratio”) conducted in Tennessee in 1986: the pupils and the teachers who took part in the experiment were assigned randomly to classes of small size (about 12 pupils) or large size (about 23 pupils), in a total sample of 325 classes. This method (cf. box 1, “randomised experiment”) makes it possible to neutralise the selection bias manifestly and transparently. Another category of methods consists in using “quasi-experiments” making it possible to cause class size to vary for reasons that are unrelated to the initial achievement or ability levels of the pupils in the various classes (cf. box 1, “quasi-experiments”). Finally, a third method consists in using the maximum class size thresholds that exist in many countries (cf. box 1, “class-opening thresholds”) as an exogenous source of variation of the number of pupils per class.

Box 2: Interpreting an effect measured in standard deviation per pupil less in the class

The effect of an education policy on the performance of the pupils is often measured on the basis of testing of the pupils. Such tests do not have a natural scale, they can be more or less difficult, they can have a larger or smaller number of questions, etc., and their absolute value has no precise meaning. It is common practice in the research literature to make them comparable by normalising them by using the standard deviation of the test in the population. Technically, this involves dividing the score obtained by each pupil by the standard deviation of the sample. Having two tests that are more or less difficult is like having two ladders with the rungs spaced apart to greater or lesser extents: normalising a test by its standard deviation is like looking at the ladder rung on which a pupil is standing, without taking account of the gaps between the rungs.

Thus, when the results are normalised by their standard deviation, what they express is more a relative level than an absolute performance. Once they are normalised, all of the tests have a standard deviation brought to 1. So, if the effect of an action is to increase the test by one (i.e. by one standard deviation, or by 100% of a standard deviation), that would be equivalent to bringing a median pupil (the 12th out of 24 pupils, for example) to the level of the pupil who is ranked 4th out of 24, i.e. progress of 4 positions in a class. An effect of 50% of a standard deviation corresponds, for the median pupil, to progress of 8 positions in the class. An effect of 10% of a standard deviation corresponds, for the median pupil, to progress of 0.8 positions in the class. An effect of 0.5% of a standard deviation corresponds, for the median pupil, to progress of 0.04 positions in the class.

Halving the class size (from 24 to 12 pupils per class) by doubling the number of classes would, according to the studies identified in this policy brief, lead to an improvement in scholastic performance lying in the range 20% to 30% of a standard deviation. Such an effect is considerable: in a class of 24 pupils, it corresponds, for instance, to progressing from one half of the mean performance gap to one quarter of the mean performance gap.

Effects on performance are established at primary level, but many unknown quantities remain

In Table 1, we present the main results of the selected studies, which cover a small sample of developed countries (USA, France, Israel, Norway, and Sweden). In this table, we look at the impact of one less pupil in the class on the mean scholastic performance levels of the pupils in that class, as measured by standardised tests (1). These tests are defined on variable and arbitrary scales (for example, the scales depend on the difficulty of the test). In order for the orders of magnitude to be comparable between the studies, it is common practice to normalise such tests by using their standard deviations in the population. Table 1 thus presents the effects of spending a year in a class with one less pupil, as expressed in a percentage of the standard deviation (box 2 explains how to interpret these measurements).

Notes: The references of each of the studies are included in the bibliography. The effects are expressed in percentage points of a standard deviation of the score in question per pupil less in a class for a school year. A positive coefficient should therefore be interpreted as a beneficial effect of the reduction in class size on pupil performance. The coefficients indicated in bold are statistically significant at the threshold of 5%. The other coefficients are non-significant at the same threshold.

Table 1: Short-term effect of class size on pupil performance in standardised skills tests

<table>
<thead>
<tr>
<th>Reference of study</th>
<th>Country</th>
<th>Impact of one less pupil on scholastic performance (in percentage of the standard deviation of the score in question)</th>
<th>Empirical method</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Elementary (Primary) School</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knugren (1999)</td>
<td>USA</td>
<td>2.06</td>
<td>Reading and maths (kindergarten)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.31</td>
<td>Reading and maths (1st Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.85</td>
<td>Reading and maths (2nd Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.44</td>
<td>Reading and maths (3rd Grade)</td>
</tr>
<tr>
<td>Hesby (2006)</td>
<td>USA</td>
<td>0.00</td>
<td>Maths (3rd Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01</td>
<td>Reading (3rd Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00</td>
<td>Writing (3rd Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00</td>
<td>Maths (4th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00</td>
<td>Reading (5th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01</td>
<td>Writing (5th Grade)</td>
</tr>
<tr>
<td>Asgrist and Levy (1999)</td>
<td>Israel</td>
<td>1.18</td>
<td>Maths (4th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.34</td>
<td>Reading (4th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.95</td>
<td>Maths (5th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.89</td>
<td>Reading (5th Grade)</td>
</tr>
<tr>
<td>Asgrist et al. (2017)</td>
<td>France</td>
<td>2.79</td>
<td>Maths (2nd Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.19</td>
<td>French (2nd Grade)</td>
</tr>
<tr>
<td>Brezoux et al. (2008)</td>
<td>France</td>
<td>2.70</td>
<td>Reading (3rd Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.60</td>
<td>Reading (4th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.20</td>
<td>Maths and French (1st Grade)</td>
</tr>
<tr>
<td>Fredrikson et al. (2013)</td>
<td>Sweden</td>
<td>3.30</td>
<td>Logical and verbal skills (4th to 6th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.45</td>
<td>Maths and reading (4th Grade)</td>
</tr>
<tr>
<td>Bjarne and Børsingrenning (2013)</td>
<td>Norway</td>
<td>0.80</td>
<td>Maths (9th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.14</td>
<td>French (9th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.63</td>
<td>History &amp; Geography (9th Grade)</td>
</tr>
<tr>
<td>Plenly and Valdamerai (2006)</td>
<td>France</td>
<td>0.01</td>
<td>Maths (6th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01</td>
<td>Maths (8th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01</td>
<td>Norwegian (8th Grade)</td>
</tr>
<tr>
<td>Leven et al. (2008)</td>
<td>Norway</td>
<td>0.00</td>
<td>Maths (9th Grade)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00</td>
<td>Swedish (9th Grade)</td>
</tr>
</tbody>
</table>

Notes: The references of each of the studies are included in the bibliography. The effects are expressed in percentage points of a standard deviation of the score in question per pupil less in a class for a school year. A positive coefficient should therefore be interpreted as a beneficial effect of the reduction in class size on pupil performance. The coefficients indicated in bold are statistically significant at the threshold of 5%. The other coefficients are non-significant at the same threshold.

Sources: Knugren (1999); table VI; normalised by the standard deviations indicated in the appendix; Asgrist and Levy (1999); table VI; Asgrist et al. (2017); table 4; Bresoux and Liman (2011); table 1; Bresoux et al. (2008); table 4.

References: knugren (1999); table VI; normalised by the standard deviations indicated in tables I and 10; Fredrikson et al. (2013); table 1; C. Gualdoni and Boreman (2013); table 3; Bresoux and Liman (2011); table 1; Bresoux et al. (2008); table 4; Hooby (2000); table IV (since the coefficients are estimated as percentages, they are converted into percentage points by estimating their values at the mean class size in the data, i.e. 21 pupils); Leven et al. (2008); table 4.
The estimated effects of class size in elementary or primary schools are generally positive: of the nine studies identified, seven found statistically significant effects. These effects generally lie in the range 2% to 3% of a standard deviation of the test per pupil less in the class, which means, for example, that halving a class of 24 pupils (i.e. a reduction of 12 pupils per class) improves the mean performance of the pupils in question by from 20% to 30% of a standard deviation (2) at the end of the year (see box 2). Those conclusions are consistent with the recent review by Monso (2014).

The second part of Table 1 presents the few studies that are available on the impact of class size at lower secondary (junior high) schools. The effects appear less marked than at primary school level. In France, for example, the class-opening thresholds method results in effects that are half those found at primary school level (3). One possibility that is often discussed is that class size measurement is less reliable at lower secondary level. Another possible interpretation lies in the differences of teaching context. However, it is very difficult to speculate on interpretation of these differences with such a limited sample of lower secondary school studies.

Beyond the effects on mean performance levels of pupils, a few studies have endeavoured to measure the effects of class size for various different sub-populations of pupils. Piketty and Valdenaire (2006), for example, obtained effects that were twice as high as those presented in Table 1 when they considered children of disadvantaged social origin (children whose parents are workers or are not in work), thus suggesting that a targeted policy of class size reduction in the priority education schools could contribute to significantly narrowing performance differences as a function of social origins.

Similarly, Krueger (1999) found higher effects for the poorer pupils and for the black pupils, and Iversen and Bonesrønning (2013) identified effects that were twice as high for children from single-parent families. However, the other studies mentioned in Table 1 either do not examine this heterogeneity of effects, or do not find significant differences relative to the mean effect. Overall, the benefit of a reduction in class size could be particularly high for the most socially disadvantaged pupils, but the literature as a whole does not give a very firm demonstration of that.

Another unknown quantity concerns the non-linearity of the effect of class size. The above-mentioned studies are, in general, based on quite large variations in class sizes, of about ten pupils relative to classes numbering from 20 to 25 pupils on average. But we do not know whether the impact per pupil more or per pupil less would be lower or higher for class size variations of larger or smaller amplitude. We are not aware of any studies capable of establishing such fine distinctions convincingly (4).

The long-term effects of class size

Beyond its short-term effects on end-of-year scholastic performance, some studies suggest that class size influences academic pathways and labour market outcomes in the longer term (Table 2).

(2) It can be noted that the order of magnitude of the effects measured by the studies cited in Table 1 is twice as high as the one obtained in the recent review by Hattie (2005), which is 13% of a standard deviation. That review incorporates a large number of studies, many of which are based neither on randomised experiments nor on quasi-experiments.

Table 2: Long-term effects of class size

<table>
<thead>
<tr>
<th>Reference</th>
<th>Country</th>
<th>Classes concerned</th>
<th>Effect of one pupil less in the class for one year</th>
<th>Empirical method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Krueger and Whitmore (2001)</td>
<td>United States</td>
<td>Primary</td>
<td>Probability of taking the college (university) entrance examination (SAT test): +0.7 ppt</td>
<td>Randomised Experiment</td>
</tr>
<tr>
<td></td>
<td>(STAR)</td>
<td></td>
<td>Scores obtained in the SAT test: +0.03 of a standard deviation of the distribution of the scores</td>
<td></td>
</tr>
<tr>
<td>Chetty et al. (2011)</td>
<td>United States</td>
<td>Primary</td>
<td>Probability of going to college (university): +0.12 ppt</td>
<td>Randomised Experiment</td>
</tr>
<tr>
<td></td>
<td>(STAR)</td>
<td></td>
<td>Salary / n.s.</td>
<td></td>
</tr>
<tr>
<td>Browning and Heinesen (2007)</td>
<td>Denmark</td>
<td>Lower Secondary</td>
<td>Number of years of studies: +0.02 years</td>
<td>Class-opening thresholds</td>
</tr>
<tr>
<td>Fredriksson et al. (2013)</td>
<td>Sweden</td>
<td>Primary</td>
<td>Number of years of studies: 0.02 years</td>
<td>Class-opening thresholds</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rate of employment: n.s.</td>
<td></td>
</tr>
<tr>
<td>Falch et al. (2017)</td>
<td>Norway</td>
<td>Lower Secondary</td>
<td>Number of years of studies: n.s.</td>
<td>Class-opening thresholds</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salaries / n.s.</td>
<td></td>
</tr>
<tr>
<td>Leuven and Lakken (2017)</td>
<td>Norway</td>
<td>Primary and Lower Secondary</td>
<td>Number of years of studies: n.s.</td>
<td>Class-opening thresholds</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salaries / n.s.</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The references of each of the studies are indicated in the bibliography. The effects indicated in column 4 correspond to the impact of one pupil less in the classes concerned by the reduction in class size (column 3). The effect not statistically significant at the threshold of 10% are indicated by the abbreviation "n.s." (not significant). The coefficients expressed in percentage points are indicated by the abbreviation "ppt".

Sources: Krueger and Whitmore (2001): section 5 and table 6 (by dividing the effect by the mean difference between the large and small class sizes, i.e. 4.4 pupils); Chetty et al. (2011): table 5 (by dividing the effects by the mean difference between the large and small class sizes, i.e. 7.5 pupils, for a time of 2.14 years); Browning and Heinesen (2007): table 2; Fredriksson et al. (2013): table IV (by dividing the effects by the number of years spent in a class of reduced size, i.e. 3 years); Falch et al. (2017): table 4; Leuven and Lakken (2017): table 3.
Based on monitoring data from the American STAR programme, Krueger and Whitmore (2001) showed that the pupils who had been randomly assigned to the small classes at primary school were more likely to take college (university) entrance exams and achieved higher marks in them than the pupils who had been assigned to the larger classes. Following on from that study, Chetty et al. (2011) showed a positive effect of having been in a small class at school on the probability of going to college (university), without however managing to detect any significant impact on salaries in adulthood. In a very different context, Fredriksson et al. (2013) show particularly spectacularly that in Sweden, people who were in smaller classes at primary/elementary school not only attain a higher level of studies, but also enjoy a higher employment rate and higher income in adult life.

There is more uncertainty about the long-term effects of class sizes at lower secondary level. In Denmark, Browning and Heinesen (2007) found that small classes at lower secondary school had positive effects on number of years of studies. Conversely, two recent articles on Norwegian data (Falch et al., 2017; Leuven et Løkkken, 2017) did not detect any significant effects of class size at lower secondary school on level of studies or on income, tying in with the absence of effect on performance in the short term shown previously by Leuven et al. (2008).

**Reduction in class size and teaching practices**

It is sometimes said that reduction in class size is effective only if the teachers adjust their teaching practices to take advantage of it, and that such a policy is therefore relevant only if it is accompanied by teacher training to that end. Such a statement has no empirical basis. The experiments and quasi-experiments used in the literature presented above do not include any specific teacher training, and yet they produce significant gains in pupil performance. In other words, empirical studies show that class size reduction is effective even if the teachers are not given specific support and assistance as regards the way they teach.

Why? This point is discussed in education science, but it would seem that even if teachers make little change in their practices when they teach in smaller classes, it is observed that the pupils are more committed to their tasks, there are fewer discipline problems, more time is spent on teaching, and there is more prevention than remediation (Meureut, 2001; Ecalle et al., 2006; Blatchford et al., 2007; Blatchford et al., 2016).

It might be legitimately thought that the effect of a reduction in class size would be even more significant if that policy were supplemented by teacher training and by tools or practices that take particular advantage of having smaller groups. But, to our knowledge, no studies exist that rigorously show such complementarity: for that purpose we would need to have an experiment or a quasi-experiment that makes it possible to establish the causal effect of a class size reduction coupled with training, and to compare it with the causal effect of a class size reduction without training.

**A costly policy that needs to be appropriately targeted**

Taken as a whole, the results of the studies identified in this policy brief indicate that a determined policy for reducing the size of classes in primary/elementary schools would significantly improve the performance of the pupils in question, and could, in the longer term, positively influence their academic pathways and their labour market outcomes.

The benefits of such a policy are however to be compared with its cost for the public purse, it being possible for that cost to be prohibitive if the class size reduction is insufficiently targeted. By taking inspiration from the figures proposed by various studies (Krueger, 2003; Fredriksson et al., 2013), the annual cost of a policy consisting in halving the size of all of the CP (Year 2/1st Grade) and CE1 (Year 3/2nd Grade) classes in France by reducing them on average from 24 to 12 pupils per class (5) can be estimated to be about 5 billion euros, i.e. an amount equivalent to over 7% of the current French National Education budget!

In addition to its high cost for the public finances, a general reduction in the size of CP and CE1 classes might give rise to a reduction in the quality of recruitment of the teachers assigned to them, as has been observed in certain American States that have implemented policies of generalised reduction in class size (Jepsen and Rivkin, 2009; Dieterle, 2015). Under these conditions, it would seem more opportune, at least in the short term, to limit the reduction in class size to primary schools having priority education status only, especially in a context in France that is already marked by recruitment difficulties (Cnesco, 2016).

Targeting the scheme on the pupils who are the most disadvantaged socially also has the aim of redistributing the narrowing of performance gaps according to social origin: based on a mean value for performance improvement of 2% of a standard deviation per pupil less in a class, the benefit of halving CP and CE1 classes would represent almost one half of the performance gap between children from disadvantaged backgrounds (with parents who are workers, or who are not in work) and children from advantaged backgrounds (with parents who have managerial or professional occupations) at the start of the CP year (see box 2). Targeting the scheme on schools with priority education status could also be more effective than a general reduction in class size if, as suggested by certain studies, the socially disadvantaged pupils benefit to a greater extent from reduced class size than their classmates from better-off backgrounds.

Based on the same calculation hypotheses as above, the cost of halving the size of all CP and CE1 classes of REP (priority education network) schools and of REP+ (priority education network plus) schools can be estimated to be a little under 700 million euros per year. The annual cost of the scheme would be reduced to 300 million euros if it were limited to REP+ schools only (6).

---

(5) This calculation is made by assuming, like Fredriksson et al. (2013), that the fixed costs induced by a class size reduction policy (building new classrooms, additional training expenses, etc.) represent 30% of the variable costs. The variable costs correspond to the wage bill that would have to be paid to recruit the teachers of the new CP and CE1 classes. In 2016, the number of CP and CE1 classes (not combined or grouped together) in public primary (elementary) schools in France can be estimated to be about 60,000 (source: Indicateurs APAE). Doubling the number of those classes (so as to halve class size) would require about 60,000 teachers to be hired, with the corresponding annual wage bill being estimated at about 50,000 euros.

(6) In 2016, there were about 6,000 CP and CE1 classes (not combined or grouped together) in the REP schools, and 4,000 classes in the REP+ schools.
Would the benefits of such a policy for the pupils in question be greater than its cost for the public finances? To try and answer this question, it is possible to calculate the salary gain that should be enjoyed, on average, by the pupils of the schools concerned by the halving in size of the CP and CE1 classes to compensate for the cost of the measure. Assuming that the mean salary of these pupils will be able to earn in their careers is close to the present mean salary of the teachers, and choosing a conventional discount rate of 5%, halving the sizes of CP and CE1 classes in schools having priority education status would be "profitable" if the salary gain for two years spent in a class whose size has been halved.

The results obtained by Fredriksson et al. (2013) on Swedish data suggest that this threshold could be exceeded, insofar as their estimations imply that a reduction in class size by 12 pupils for two years at primary (elementary) school would increase the salaries of the pupils in question by 5% on average.

**Conclusion**

While certain authors think that it is pointless to assign additional resources to the education system and recommend systemic reforms (such as introducing competition between schools), recent research on class size reduction makes it possible, on the contrary, to justify a determined policy of public spending on education. The soundest studies make it possible today to consider that the effects of class size reduction are relatively high, going against an idea that persisted for a long time in the education world. The cost-benefit calculation that we have proposed would suggest that the budgetary effort imposed by such a policy is very likely a profitable investment for the future, especially if it is appropriately targeted.

Another approach to education policy that seems less costly and more consensual consists in changing the teaching practices implemented, through training actions that follow the recommendations from the research. The order of magnitude obtained for a reduction in class size of about twelve pupils, i.e. 20% to 30% of a score standard deviation, is, admittedly less than the effects of teaching actions conducted in extremely controlled contexts, such effects easily approaching 50% to 100% of a standard deviation.

But the actions or policies that are implemented on a large scale, or in "ecological" situations, and that are rigorously assessed, rarely present such high effects. For example, in France, highly structured teacher training on preparing for reading have effects assessed at 15% of a standard deviation (Bouguen, 2016), while an action that is implemented outside the class, that is very intensive, and that is implemented in small groups, such as the "Coup de pouce clé" ("Key helping hand") scheme, does not have any demonstrated added value (Goux et al. 2017). In the United States, abundant literature estimates the effect of Charter Schools: most of those schools have no impact; only the best of them, those that adopt a specific teaching approach, have demonstrated effects, but they are then comparable with the effects of halving class size. Along similar lines, the Internat d’Excellence de Sourdun (Boarding School of Excellence of Sourdun) also has mean effects of 20% of a standard deviation, for a cost per pupil that is comparable to that of halving class size (Behaghel et al., 2017).

There are too few rigorously assessed education policies, in particular in France, for it to be possible to perform systematic comparison, but it remains to be proved whether many other actions are available whose demonstrated effects, on a large scale, would be higher. Actually, many studies teach us that training often fails to change the practices of teachers in depth, through a lack of intensity and of sufficient support (8). Changing practices is not something that can be decreed. Admittedly, the teacher training should be pursued and reinforced but, by contrast, a class size reduction policy can be decided and implemented by the public authorities in quite a definite manner if it devotes the necessary resources to it. This is what gives it its force and its advantageousness. The literature shows that this impulse suffices to produce significant effects.

This review also shows that our knowledge is actually limited to a few orders of magnitude, while numerous details – are there any non-linear effects? To what extent do the effects differ according to sub-population? What role could teaching support play? What differences are there between the scholastic achievement levels? What context effects are more favourable? etc. – remain unexplored in the context of very robust methods. It is surprising that, on a topic as important and as debated, there has been so little experimental research making it possible to understand more clearly what is at stake.

**Authors:** Adrien Bouguen is a post-doctoral student at the University of Mannheim. Julien Grenet is a researcher at CNRS, an associate professor at the Paris School of Economics, and Assistant Director of the Institut des Politiques Publiques. Marc Gurgand is a researcher at CNRS, associate chair at the Paris School of Economics, and scientific director of J-PAL Europe.

---

(8) See Jacob (2017) for a recent example and illustrations.
Bibliographic References


Cour des comptes (2016) Rapport public annuel, La Documentation française.


