Teachers' Training, Class Size and Students' Outcomes: Evidence from Third Grade Classes in France[1]

Pascal Bressoux*, Francis Kramarz** and Corinne Prost***

* Université de Grenoble

** CREST, CEPR, IZA

*** CREST, EHESS Paris-Jourdan, Cornell University

Preliminary

December 2005

Abstract

This paper studies the impact of different teacher and class characteristics on third graders' outcomes. It uses a feature of the French system in which some novice teachers start their jobs before receiving any training. Three categories of teachers are included in the sample: experienced teachers, trained novice teachers and untrained novice teachers. We find that trained and untrained novice teachers are assigned to similar classes, whereas experienced teachers have better students located in better environments. Hence, in order to match similar students and classes, we focus on pupils with novice teachers and discard those with experienced teachers. In addition, we show that the same sample can be used to estimate the causal effect of class size on students' outcomes. Our findings are: (1) teachers' training substantially improves students' test scores in mathematics; on reading scores, teachers' training is beneficial only to students in high achieving classes; (2) teachers' education background has a significant impact since untrained teachers who majored in sciences at university compensate for their lack of training, they have the same effect as trained teachers; (3) the effect of class size is substantial and significant, a smaller class size improves similarly all students' reading test scores within a class and is more beneficial to less achieving students in mathematics; all students in less achieving classes are much more sensitive to class size than students in more able classes.

[1] We have benefited from helpful comments by participants in the labor economics seminar at Cornell University and the seminar of the Department of Evaluation of the Ministry of Education (Direction de l'Evaluation et de la Prospective). We are particularly grateful to Ronald Ehrenberg, George Jakubson and Robert Hutchens for their suggestions on a previous version.

Introduction

The literature on the effects of class size on student learning is huge. Yet there is no consensus on the impact of class size and the debate is still impassioned[2]. Some economists, who do not believe much that smaller class size can improve students' performance, or who find that it is a very costly policy, argue that other policies besides class size reduction, such as improving teacher quality, are more important.

Understanding the relationship between teachers' characteristics and students' achievement is obviously of prime importance in the analysis of the education system. Research on this topic has often focused on specific characteristics such as teachers' diplomas, experience and salaries. Few studies have specified the impact of teacher in-service training in developed countries. Angrist and Lavy (2001) present an evaluation of the effect of in-service teacher training in Jerusalem schools. They find that the causal effect of the program on pupils' test scores is significantly positive. The cost-effectiveness analysis suggests that teacher training may provide a less costly means of improving pupil achievement scores than reducing class size or adding school hours.

In France, most studies on teachers have looked at teaching practices, and little empirical work has examined the consequences of teachers' training on students' outcomes. Bressoux (1996) partly fills this gap. In order to study the effect of teachers' training and experience on third-grade pupil achievement, he uses a specific survey on third grade students and teachers in 1991, with a quasi-experimental design. This data source includes three types of teachers: untrained novice teachers, trained novice teachers, experienced teachers. Bressoux finds that training improves students' scores in mathematics. Experience seems also to have a positive impact on pupil achievement.

[2] For a comprehensive survey on the topic of class size effect, see for instance Ehrenberg, Brewer, Gamoran and Willms (2001)

Importantly though, the experiment used in the above study is not randomized. The ideal situation would involve the random assignment of pupils to the different types of teachers. In fact, Bressoux (1996) shows that classes differ according to the status, experienced, trained novice or untrained novice, of the teacher. Hence, in the absence of random assignment, Bressoux estimates the impact of training using regressions controlling for numerous variables. The estimated effect is the causal one if no unobserved student or class characteristic is correlated with the teacher's type and with the student's test scores. Otherwise, estimates are potentially biased.

This paper uses the same data, but relies on a methodology that takes care of the non-randomized design. The idea comes from the specificity of experienced teachers. The fact that the allocation of classes is not random is virtually only due to experienced teachers, who can choose their schools, and who often choose advantaged zones. But, in principle and in the data, trained and untrained novice teachers are assigned to almost similar classes. So our paper uses the fact that, when excluding experienced teachers, we are faced with a quasi-randomized design. We check the robustness of this feature using different estimation methods, both conditional and unconditional on other observed variables.

The data used here are very rich. The unit of observation is the student, a very important element for this kind of analysis (see Summers and Wolfe, 1977). Multiple students' characteristics are collected. Furthermore, all students within a third-grade class are included in the sample. This gives us an opportunity to control for class effects. In addition, teachers also provide a lot of information on their personal characteristics, their teaching practices, as well as characteristics of their classes and their schools. Moreover, students' achievement is extremely precisely measured by detailed test scores at the beginning and at the end of the year.

A first aim of this paper is to check that Bressoux's findings on training – better trained teachers induce higher students' outcomes – are robust. To perform this task, we use more recent statistical methods, controlling for the endogenous allocation of classes. The estimation is made excluding experienced teachers, in order to estimate the causal effect of training of novice teachers on pupils. Particular attention is given to heterogeneous effects. A second goal is to see if some particular characteristics of the teachers, such as their university background (which was not used in Bressoux (1996)), have any impact on their students' outcomes. This paper also examines other class characteristics, more particularly class size. Indeed, when excluding experienced teachers, it appears that class size is not correlated with pupils' initial test scores. There is no sign of a relation between class sizes and class mean initial achievement or class socio-economic background. Thus, it seems that no selection bias in class size allocation is present when the sample is restricted to novice teachers. Consequently, we use similar methods to assess the effect of class size as were used to estimate the effect of training effect.

The findings on the training effect are very close to those found by Bressoux (1996): the training of novice teachers promotes students' learning in mathematics. Yet it seems that within classes, less able students do not benefit from their teachers' training. In addition, training allows teachers to improve significantly students' reading scores only in high achieving classes.

We also find that teachers' education background has a significant impact since untrained teachers who majored in sciences at university have the same effect as trained teachers. It seems that their past studies help them to compensate for the lack of training.

The estimated impact of class size implies that reducing class size has a positive and substantial effect on third-graders. These results are close to the findings of Piketty (2004) on the effect of the size of French second-grade classes. It appears that the effect is similar on students' reading scores within the classes; it is larger for less able pupils in mathematics. Moreover, a smaller class size improves more students' scores when they are in a less achieving class, which could be the consequence of higher frequencies of disruptions in this kind of classes, as described in Lazear (2001).

The paper is organized as follows. Following a description of the data in Section I, Section II describes the statistical model and the empirical tests. Section III reports the main estimation results and Section IV concludes.

I. Data and descriptive statistics

The data come from a survey conducted by the French Ministry of Education. They cover a sample of classes of third-graders (8 years old) and their teachers. The quasi-experimental design is due to a feature of the system of teachers' training in France. This characteristic implies that some novice teachers start their job before any training.

In France, except for a subset of private schools, teachers are civil-servants recruited and paid by the State. After having passed a competitive examination, primary school teachers are trained in specific schools. At the beginning of the 1990's, these schools were called 'écoles normales'. France was, and still is, geographically divided into administrative 'départements' and there was an 'école normale' in each 'département'.[3]

Novice teachers are recruited among students who have passed a competitive examination for entering an 'école normale'. To take this examination, students have to have already passed an examination corresponding to two years in a university. The number of slots in the école normale' is limited and determined each year at the central level, using forecasts for teachers' positions. All applicants are ranked according to their grades in this examination. The students ranked first enter the 'école normale' and are trained during two years. Students who are ranked just after the last admitted candidate on this primary list are assigned and ranked within a waiting list.

In September, the number of vacant job slots is often greater than the one expected two years earlier. So students who have finished their training at the 'école normale' are assigned to some of these job slots, and, in October, some students in the waiting list are

[3] The training schools are now called university teacher training institutes 'Instituts Universitaires de Formation des Maitres (IUFM)' and belong to a region (a region includes several departments).

assigned to the vacant slots. Hence, these persons have to teach a class for an entire school year without receiving any training. They enter the 'école normale' the year after.

The survey was conducted in the school year 1991-1992. The sample included explicitly three categories of teachers: untrained novice teachers, trained novice teachers and experienced teachers. The sample covered third-grade students and their teachers in 12 'départements'. The teachers were teaching in third grade classes or in multi-grade classes including third graders. In the 12 'départements' selected, all novice teachers were surveyed while experienced teachers were chosen randomly. Finally, the survey covered 4,001 students and 197 teachers. The numbers of teachers within each category were not perfectly balanced: there were 96 experienced teachers, 65 trained novice teachers and 36 untrained novice teachers (see table 1).[4]

The information about the students is comprehensive: parents' occupations, sex, month of birth, nationality (French or not), number of siblings, number of years spent in a pre-elementary school, repeated classes (see the statistics in table 2). In addition, two sets of scores are available in the data. In France, there is national testing of all pupils just at the beginning of the third grade, both in reading and mathematics. The reading tests comprise grammar, vocabulary, spelling and reading comprehension *per se*. The mathematics tests comprise arithmetic, geometry and problem-solving. For this specific survey, covered pupils have also been tested at the end of the school year in both subjects, using a design similar to that prevailing in the entry tests. For each of the two subjects, initial and final scores are standardized (mean=100, standard error=15).

In addition, teachers had to answer a questionnaire on their personal characteristics, on their teaching practices and on the characteristics of their classes and their schools. The main variables used in the following are the field of specialization of the teacher during his/her studies at the university (sciences, unknown, other), the class size, the fact that the class is or not a combination class mixing students across grades, the category of the area

[4] These statistics are slightly different from the ones in Bressoux (1996) because the matching of student data and teacher data has been made a bit differently.

of the school (rural, semi-rural, urban), and the priority status (see the statistics in table 3). The mean of the class sizes is 23.9 students per class, with a standard deviation of 4.1. By comparison with data on all elementary schools, Piketty (2004) finds that the average class size in the primary schools (first grade to fifth grade) is close to 23.3 in the school year 1991-1992.

Unfortunately, the scores are not available for all the students. This attrition comes from two reasons. First, some students were not in class when the tests were conducted. Second, for some classes, all the scores are missing. The scores of reading tests are not known for 974 students and the scores of math tests are not known for 778 students. The class size also is not known for all classes: for 8 classes, the class size is unknown and can not be approximated by the number of students of the sample, because these classes mix students of different grades.

Nevertheless, this attrition should not induce any bias: tables 4 and 5, compared to tables 2 and 3, show that the characteristics of the students whose scores and class size are known do not significantly differ from those of all the students. It seems that the absence of information on the scores or the class sizes have random origins.

The data do not come from an experimental design. In fact, the assignment of the different types of teachers to the classes is not randomized. Indeed, the system of job assignment depends on the teachers' choices. When the choices of different teachers are the same, the final assignments depend on the years of experience and on a mark given by the administration, this mark being well correlated with the years of teaching experience. Hence, as they accumulate experience, teachers are able to choose the schools they want, and mostly go from disadvantaged schools to advantaged ones. On the contrary, novice teachers go to schools that have not been chosen by experienced teachers, or where there are free job slots because some experienced teachers retired or are absent for the year.

The data show that the aggregate characteristics of pupils vary with teachers' types (see tables 6 and 7). Indeed, experienced teachers have on average better classes. In these classes, compared to those with novice teachers, initial scores are higher, the share of non-French pupils is lower, children have fewer siblings, fathers and mothers have more often a high occupation and students less often repeated the first grade. In addition, the class sizes are on average larger, and the schools are less often in a priority educational area.

The classes with trained novice teachers and those with untrained novice teachers are more alike. Nevertheless, trained novice teachers are more often in urban areas and in priority zones than untrained novice teachers.

There is a potential source of bias due to the fact that the trained novice teachers may have had better rankings at the entrance examination at the 'école normale' than the untrained novice teachers. If the examination measures the initial teaching abilities (a fact that should be proved), this bias could imply that the trained novice teachers are better able to teach than the untrained novice teachers. Fortunately (for us), the survey has been conducted during the school year 1991-1992, which is an atypical year, as can be seen in figure 1. Indeed, in 1991, the number of students selected for entry into the teacher training centers was very small. So the surveyed untrained novice teachers, who had taken the entrance examination in 1991, had very good rankings and would have been selected for entry had they competed for the examination during another year, and especially during the year 1989, when the surveyed trained novice teachers had passed their entrance examination. So the selection bias is likely to be weak.

II. Statistical method

The non-randomized assignment of the three types of teachers can also be observed through a regression of initial test scores on student and teacher characteristics. If the coefficients of the dummy variables for the types of the teacher are significant, it means that the assignment is non-randomized since the students have not been exposed to these teachers' teaching yet.

The regression is estimated on all the students. It includes random class effects, in order to take account of the correlation between students within classes. Indeed, class variables may be not sufficient to control for these correlations. So, it is important to incorporate class effects: without them, the standard deviations could be underestimated. It would be the case with OLS estimation (see Moulton (1986)). However, Moulton stresses the problem of the precision of coefficient estimates, but he also shows that the coefficients may be different when the estimation incorporates random class effects without imposing the absence of correlation between these effects and the other covariates. Indeed, this kind of estimation results in substantial gain in efficiency. Throughout this paper, class effects are estimated through mixed models (see Robinson (1991)). These models allow a general specification of class effects, fixed effects being only a specific case of this specification. Identification of class effects uses more information than for "classic" fixed effects: it uses the variance of the class effects instead of only the mean, thanks to a more general prior distribution (see appendix A).

The results of the regression of initial test scores on teacher type are detailed in table 8 (full results in table 11). They confirm that experienced teachers teach in better classes. Table 8 reports that the correlation between student initial scores and the dummy variable for the teacher experience is significant, in reading and in mathematics as well. These two correlations remain significant, even when controlling for student characteristics. On the contrary, it seems that classes with untrained novice teachers and classes with trained novice teachers are not different in terms of initial achievement, since the correlations between initial scores and the dummy of the teacher training are not significant, with or without other controls. Thus it seems that there is no selection of trained teachers, so that the classes of such teachers appear similar to those of untrained teachers according to pupils' initial achievement.

This idea is checked with the same regression of initial scores, but with the sub-sample of the pupils having novice teachers. The results are given in table 9 (full results in table 12). The coefficient of the training dummy is never significant.

Bressoux (1996) assumes that the selection bias of experienced teachers can be controlled with the observed variables, including initial test scores. The causal interpretation of the coefficients related to the type of teacher relies on the assumption that no selection bias comes from unobserved variables.

This paper takes care of the non-randomized design in order to assess the robustness of the teachers' training effect found in Bressoux (1996). The idea is that trained and untrained novice teachers are randomly assigned to classes, at least according to our observed variables. Hence we have chosen to estimate the training effect on the sub-sample of novice teachers. In the spirit of matching classes to classes, either taught by trained or untrained novice teachers, we focus on pupils with novice teachers and discard those with experienced teachers. It means that we manage to have a sample of similar students, some have trained novice teachers and constitute the treatment group, and some have untrained novice teachers and constitute the control group. Thus we can expect that no bias perturbs the coefficients in the estimation and that the coefficient of the treatment estimates the causal effect. The idea is close to the one in Angrist and Lavy (2001). In this paper, they observe that pupils in the treatment group have initially lower score than pupils in the control group. As they would like pupils in the control group to be comparable to pupils in the treatment group, they match individual pupils on the basis of their initial test scores, by dividing test scores into quartiles and comparing treatment and control scores in each quartile. Here, we restrict the sample in order to have similar pupils in the treatment and control group. But, on the contrary to Angrist and Lavy, we keep a regression strategy, in order to control for the other covariates, and more specifically to control for class effects. We will see that these controls are important.

Thus, we will be able to estimate the effect of training on achievement using this specific sample. Nevertheless, we will have to keep in mind this restriction while interpreting the results: what we estimate is the effect of a trained novice teacher on pupils' achievement, compared to the effect of an untrained novice teacher, and not compared to all other kinds of teachers.

In the meantime, we will keep this strategy to estimate the class size effect. Indeed, table 8 reports that the correlation between initial scores in reading and class size is positive and significant when all pupils are included in the regression. When adding other covariates, this correlation remains significant and positive, even if it is less significant. On the contrary, table 9 reports that even without any other control, class size is no more correlated with initial scores when the sample is reduced to the students with novice teachers.

Figures 2 and 3 present these results. These figures show the link between class sizes and class means of initial test scores in reading. The classes with experienced teachers are presented in figure 2 while figure 3 presents the classes with novice teachers. It is clear that all experience teachers teach high achieving classes whereas the scores of the classes with novice teachers are much more dispersed. Also, experienced teachers more often have larger classes. At last, the positive correlation between class size and scores can be seen in figure 2 with experienced teachers, even if it is not obvious, and it appears in figure 3 that there is no more correlation for those classes with novice teachers.

The idea that class size can be positively correlated with student achievement is well known: the education system is often organized in order to support less advantaged pupils by gathering them in small classes whereas more advantaged students are assigned to larger classes. Hence the differences in class sizes are often in relation to students' socioeconomic background and scores. The selection bias in the relationship between test scores and class size can be generated within schools as well as between schools. This selection bias is one reason why causal effects of class size can be difficult to measure.

There are several reasons explaining why, in France, the selection could be weak for third grade classes. First, the system of assignment of teachers is centralized, and is not supposed to make any difference between schools in terms of resources. The only official exception is the policy of education priority areas (ZEP, 'zones d'éducation prioritaires'). The ZEP policy is a program implemented in 1982, which gives more resources to disadvantaged schools (for a description and an assessment of this program on sixth and seventh graders, see Benabou, Kramarz, Prost (2003)). According to our data, the classes in the ZEP have on average 23.8 students per class, whereas the mean class size in the non priority zones is 25.2[5].

The other case where it is known that some schools have smaller classes than the other schools is the one of rural schools: because of small enrollments, these schools have often small classes, even if they often organize combination classes by mixing students of different grades in a class. Yet the conclusion in terms of selection is not clear since, as we will comment on this later, pupils in rural schools have better achievement at the beginning of the third grade (but improve less during the year).

Nevertheless, there may be selection within schools. This selection is possible in large schools, when there are several third grade classes. Yet we will see that when the enrollment exceeds 30 students, it does not always entail a new third grade class, but sometimes some third graders are assigned to a class with students of other grades. To facilitate this assignment, the school may choose good pupils to go to this combination class, so that students who stay in the larger class are not necessarily the better ones.

The organization of a selection needs large schools. Since experienced teachers are much more often in urban areas, where schools are bigger, this may explain why the selection on initial scores can be observed for classes with experienced teachers and not for novice teachers. On the contrary, it seems that pupils with novice teachers are not assigned to classes with different sizes according to their abilities.

[5] The priority zones are more often in urban areas, where classes are larger than in rural areas. So the effective reduction in class size in ZEP schools could be larger than the one given by the raw difference of the two means. A regression of the class size on the dummy variable for ZEP schools, controlling for the rural areas and the combination classes, give a class size smaller of 1.7 students in priority zones.

Finally, we will estimate the class effect on the sample excluding experienced teachers. As the correlation between class size and observed initial scores is significant on the whole sample, we suspect that there may also be a selection on unobserved variables, which could disturb the estimation of the causal effect of class size. On the contrary, the correlation between class size and initial scores is not significant with the sub-sample of pupils with novice teachers. Hence we assume that the “traditional” bias selection is expurgated. Finally, to check the robustness of our findings, we will also estimate the class size effect on all the students, using instrumental variables.

III. Results

A. Global effects

The results of the estimation of the effects of teacher and class characteristics on pupil achievement are reported in table 10 (details in table 13). It is a regression of final scores on initial scores and student, teacher and class characteristics. The estimation includes class effects and is estimated on the sub-sample of students with novice teachers.

The data include a lot of information about the teachers and their teaching practices. They include in particular the diploma, the subject studied at university, the number of hours per week used for teaching reading or mathematics, the number of hours asked for homework per week, the practice of organizing the class in groups, and how these groups are chosen.

All these variables have been tested in the regression of final test scores on initial scores, student and class characteristics. When the variables on the number of hours per week used for teaching reading or mathematics, the number of hours asked for homework per week, the practice of organizing the class in groups, and how these groups are chosen are added, the coefficients on these variables are not significant. The small number of classes in the sample may prevent from identifying these effects that may be non linear. Therefore, these variables are not included in the final estimation.

The only teacher characteristic that is used is the subject studied at university. More precisely, dummy variables are included for teachers having majored in sciences at university (14% of novice teachers) and for teachers having majored in a discipline not reported in the survey (roughly 14% of novice teachers). The reference group therefore comprises those teachers who majored in humanities (often French or another language, sociology, psychology, history). Novice teachers are all endowed with similar diplomas since it is compulsory to have a diploma equivalent to two years university to enter an 'école normale'. This was not the case in the past, and among experienced teachers, only a few went to university.

The regressions of final test scores include class characteristics. Some class characteristics can be calculated using the means of individual characteristics. We built class variables such as the share in the class of students with advantaged parents as measured by occupations, the share of girls, the share of non-French students and the share of students who repeated at least one grade. These variables are calculated for each student, excluding his/her own characteristics in the calculation of the means. None of these variables give significant coefficients. They are not included in the regressions presented in this paper. This confirms the difficulty in estimating peer effects without a clean experimental design.

On the contrary, means and standard deviations of initial test scores per class have significant correlations with final scores. For the regression of final scores in reading, the included variables are the class means and standard deviation of initial scores in reading. Likewise, the means and standard deviation included in the regression of scores in mathematics are calculated on initial scores in mathematics. These means are also calculated for each student, excluding his/her own characteristics. Table 13 reports the effects of class characteristics on final test scores, and means and standard deviations of initial test scores have negative impacts on pupil improvement, meaning that students have better results in a homogeneous class and when the average achievement is not too high.

The estimated impact of training is not significant on reading achievement but it is significant and large on mathematics achievement: students gain more than 3 points on their final scores when their teachers have been trained. This effect is substantial; it is more than one fifth of the standard deviation of final scores. These results are close to the findings in Bressoux (1996).[6] They are also close to the raw differences of the means: as can be seen in table 6, students with untrained novice teachers have similar initial scores than students with trained novice teachers; yet, they improve much less during the year. The raw differences-in-differences estimator gives an effect of 2.5 in reading, and 2.1 in mathematics. Incorporating student and class characteristics show that the effect is larger in mathematics, since it is close to 3.6. On the contrary, the effect is much weaker in reading, since it is close to 1.3 and is not significant. The estimation of the regression without class effects would have drawn to a significant effect equal to 2.6. Hence incorporating random class effects shows that the coefficient is weaker and non significant.

The teachers' educational background has also a substantial impact. The finding is that teachers who majored in sciences improve their pupils' mathematics achievements more than other teachers do. These teachers are either trained or untrained. The effect is not significantly different for these two kinds of teachers, and not significantly different from the training effect. Hence, even though the training effect is substantial in mathematics achievements, teachers who have not been trained, but who have studied mathematics or sciences when they were at university, compensate for this lack of training. Nevertheless, the sample is small and this result is weak.

The teachers whose fields of specialization are unknown seem to improve their students' achievements, the effect being very significant in reading and slightly less significant in

[6] In Bressoux (1996), the effect of teacher training is estimated at 0.72, non significant, on reading scores, and 3.37, significant, on mathematics scores.

mathematics. This group of teachers comprises some individuals who did not report this information, potentially because of multiple fields of specialization, as well as some teachers who did not go to university. Indeed, very few people were entitled to take the examination for entering an 'école normale' without having studied at university; this was particularly the case of mothers of three or more children.

Nevertheless, it is worth noting that the repartition of teacher among classes according to their field of specialization is not random. The correlations between initial test scores and teacher dummy variable for unknown field of specialization are significant for the initial mathematic achievement (table 12). Hence, even if the regression of final score control for initial scores, estimates may be affected by selection bias since these teachers appear to be assigned to better classes.

Class size has also a significant impact on students' outcomes. The impact is quite similar in reading and in mathematics. For test scores in reading and in mathematics, the estimated effect is -0.34. This impact is substantial: reducing the class size by 10 students increases the final test scores by 3.4 percentage points. This is nearly the same impact as the one obtained for teachers' training in mathematics.

This impact seems to be robust to the problem of combination classes. Indeed, the regression is estimated with a sample including multiple-grade classes. In the case of a combination class, the class size is then the size of the entire class, and not the number of third-graders. Yet the dummy for multiple-grade classes is not significant[7]. Likewise, results are similar when excluding these combination classes.

The coefficients of the other class characteristics are also of interest. Students in rural schools increase their achievements much less than the other students. However, as can be seen in table 12, their initial scores are higher. These results are consistent with Brizard (1995) and Thaurel-Richard (1995): pupils in rural schools have better scores at the beginning of their third-grade, but then they tend to improve less.

[7] This result is consistent with Oeuvrard (1995).

Students in ZEP schools improve less during the year than other students do. Nevertheless, they have lower scores even at the beginning of the school year. So it is difficult to conclude that it is a causal effect.

B. Heterogeneous effects

We then estimate heterogeneous training and class size effects. To estimate the effects on heterogeneous students, the methodology chosen is to split training and class size variables into different kinds of students, in the same regression. Indeed, it is important to keep controlling for class effects. This methodology allows us to estimate the effect on some kinds of students within the classes. In addition, we estimate the effects on heterogeneous classes. Thus we will be able to interpret more precisely some results on heterogeneous effects. Indeed, some papers find that the class size effect is larger for low achieving students. We will see that this can be explain by the fact that these low achieving students are more often in disadvantaged classes; but class size affects similarly all students within a class.

Tables 14 and 15 present the results for the regression of final test scores including heterogeneous effects. These effects are measured by breaking down the dummy variable for trained teachers or the class size variable according to the quartiles.

To estimate the heterogeneous effects on pupils, the quartiles are defined by students' initial scores in reading and are measured within the classes. No significant training effects appear for scores in reading. For scores in mathematics, the training effect is substantial and significant for the more achieving students. On the contrary, it is not significant for the low achieving students. It seems that training helps teachers to improve students' results in mathematics, except for the least able ones.

In order to identify different kinds of classes, the quartiles are estimated on the class means of initial scores in reading. The decomposition reveals significant effects of training on scores in reading for the highest achieving classes. In mathematics, the effect seems to be significant in more achieving classes and not in low achieving classes. Yet the Fisher tests accept the hypotheses that the coefficients for the third and the fourth quartiles are not significantly different from the coefficient for the first quartile. So training seems to help teachers to improve their teaching, except when they face a class where the mean achievement is low: training is no help for less advantaged classes.

Among students, no heterogeneity of the class size effect appears for scores in reading. So it seems that within the classes, class size affects similarly all students. Yet, the effect on mathematics scores decrease when the “quality” of the students increases.

Among classes, the class size effect appears much more substantial for less advantaged classes and decreases when the “quality” of the class increases. The clearer results are for the scores in mathematics; the Fisher tests accept the equality of the coefficients for reading tests.

This heterogeneity is confirmed by table 16, which reports the results of the estimation of class size effect for all schools and for ZEP schools alone, estimated in the same regression. The coefficients are significant only for ZEP schools and are very substantial: -0.6 in reading and -1.1 in mathematics. This finding confirms recent results obtained by Piketty (2004) who also finds substantial impacts of class size in ZEP school, albeit marginally significant because of the small number of students in ZEP schools in his sample.

These results show that the students in ZEP schools and in disadvantaged classes in general are more sensitive to class size as a group than the other groups of students. It may come from problems of behavior in class, the probability of a troublemaker among students of a class being larger in these schools (see Lazear, 2001).

C. Instrumental variables for the class size effect

The effect of third-grade class size, as estimated in this paper, stands between -0.3 and -0.4 percentage point of final test scores. Piketty (2004), on second-grade class size, finds an impact of -0.4 to -0.5 percentage point. He applies a methodology developed in Angrist and Lavy (1999). His method is based on the following specificity of French class openings: when second-grade enrollment goes beyond 30, another class is opened in most cases. Hence, the two new classes have an average size of 15 pupils. Piketty uses this discontinuity as an instrumental variable. He finds that a reduction in class size would induce a significant and substantial increase in mathematics and reading scores, and that the effect is larger for low achieving students.

In our data, we find similar specificities as those observed by Piketty (2004) (see figure 4).[8] There are often two classes when the number of third-graders in the school is greater than 30. Yet there are some classes gathering up to 34 pupils. And the link between class size and enrollment is complicated by the existence of combination classes. When the enrollment goes just beyond 30 students, the schools do not open another third-grade class, but instead, assign some third-graders to classes with students of other grades.

When we exclude combination classes, there is less diversity in class sizes. Figure 5 shows the link between the enrollment of third graders and the class sizes. Yet when there are two classes, these classes have often different sizes; it could then be a source of bias if the sizes are determined according to the socio-economic background or the achievement of the students.

To check the robustness of our class size effects estimated on the novice teachers, we use instrumental variables on the whole sample. The instrument is based upon the enrollment of third-graders in the school when we exclude combination classes. In order to work with all classes, the instrument is also based upon the numbers of third-graders and

[8] On the contrary to Piketty (2004), Figure 4 shows all classes, including combination classes. In addition, the classes are third-grade classes and not second-grade classes. At last, our data are less reliable than those used by Piketty because we do not always observe all third-grade classes in schools.

students who are in a class with third-graders (see figure 6). In all cases, the instrumental variable is the mean of the class sizes in the school: the sizes of the third-grade classes in the first case and those of all classes with third-graders in the second case. This instrument takes care of the selection bias which exists when schools organize classes so that small classes gather low achieving students and high achieving students are assigned to larger classes.

As can be seen on figures 5 and 6, the instrumental variable is very close to the actual class size. Indeed, in our data, we identify few schools with more than one third-grade class. And when there are two classes, the sizes of these two classes are not very different. Hence the findings are easy to foresee: the results estimated with the instrumental variable are very close to the OLS results.

The idea in Angrist and Lavy (1999) is to use the discontinuity of the class size resulting from the creation of several classes when the enrollment goes beyond some level, assuming that this discontinuity is exogenous. One way of using this discontinuity is to estimate the class effect only when the enrollment is close to the “breaking point”. We use this method by estimating our instrumented regression for school where the enrollment is close to 34 students, the “breaking point” according to our data. We have chosen to restrict the sample to enrollments between 29 and 40 or between 24 and 45. The coefficients are then much more substantial, even if they are not always significant (see table 19).

All these results confirm the size of the effect: class size effect is between -0.3 and -0.5 percentage point of the final test scores.

IV. Conclusions

Thanks to the use of other statistical methods, this paper confirms the finding of teachers' training effect found in Bressoux (1996). The data used have a quasi-experimental design; the French system is such that some novice teachers teach before being trained. The effect of teachers' training is substantial: final test scores in mathematics of students with a trained teacher are greater by 3 percentage points than the scores they would have had if their teachers had not been trained. The estimation of heterogeneous effects shows that the training effect on reading achievement is significant in high achieving classes.

The importance of teachers' training is confirmed by the effect of teachers' educational background. Teachers who majored in sciences at university improve their students' outcomes in mathematics. This impact is the same for trained and untrained teachers. It means that for the untrained, past scientific studies compensate for the lack of training in mathematics.

The effect of class size is shown to be significant and negative: a smaller class size improves student achievement. The impact is evaluated between -0.3 and -0.5 percentage points. Hence, training teachers is equivalent to reducing class size by 10 students, in terms of final test scores in mathematics. It is worth noting that this equivalence is true on average. But the effects vary according to the characteristics of the classes. The effect of class size is even more beneficial in low achieving classes; these students would benefit most from a decrease in class size. The effect is particularly large for classes in priority education areas. On the contrary, it seems that this type of classes do not benefit from the training of their teachers.

References

Angrist Joshua D., Lavy Victor (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", Quarterly Journal of Economics, Vol. 114, No 2, pp. 533-574

Angrist Joshua D., Lavy Victor (2001), "Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools", Journal of Labor Economics, Vol. 19, No 2, pp. 343-369

Benabou Roland, Kramarz Francis, Prost Corinne (2003), "Zones d'éducation prioritaire : quels moyens pour quels résultats ? Une évaluation sur la période 1982-1992", CREST Working Paper, No 38

Bressoux Pascal (1996), "The Effects of Teachers' Training on Pupils' Achievement: the Case of Elementary Schools in France", School Effectiveness and School Improvement, Vol. 7, No 3, pp. 252-279

Brizard Agnes (1995), "Écoles rurales, écoles urbaines : performance des élèves en français et en mathématiques", Éducation et Formations, Vol. 43, pp.105-111

Ehrenberg Ronald G., Brewer Dominic J., Gamoran Adam, Willms Douglas J. (2001), "Class size and student achievement", Psychological Science in the Public Interest, Vol. 2, No 1, pp. 1-30

Hanushek Eric A. (1997), "Assessing the Effects of School Resources on Student Performance: an Update", Educational Evaluation and Policy Analysis, Vol. 19, No 2, pp. 141-164

Hanushek Eric A., John F. Kain, Steven G. Rivkin (2005), "Teachers, Schools, and Academic Achievement", Econometrica, forthcoming

Hoxby Caroline M. (2000), "The Effects of Class Size on Student Achievement: New Evidence from Population Variation", Quarterly Journal of Economics, Vol. 115, No 4, pp. 1239-1285

Krueger Alan B. (1999), "Experimental Estimates of Educational Production Functions", Quarterly Journal of Economics, Vol. 114, No 2, pp. 497-532

Krueger Alan B. (2000), "Economic Considerations and Class Size", Economic Journal, Vol. 113, pp. 34-63

Lazear Edward P. (2001), "Educational Production", Quarterly Journal of Economics, Vol. 116, No 3, pp. 777-803

Moulton Brent R. (1986), “Random Group Effects and the Precision of Regression Estimates”, Journal of Econometrics, Vol. 32, No 3, pp.385-397

Oeuvrard Françoise (1995), “Les performances en français et en mathematiques des écoles à classe unique”, Éducation et Formations, Vol. 43, pp.113-116

Piketty Thomas (2004), “L’impact de la taille des classes et de la ségrégation sociale sur la réussite scolaire dans les écoles françaises : une estimation à partir du panel primaire 1997”, mimeo

Robinson Geoff K. (1991), “That BLUP is a Good Thing: the Estimation of Random Effects”, Statistical Science, Vol. 6, Issue 1, pp. 15-32

Summers Anita, Wolfe Barbara (1977), “Do Schools Make a Difference?”, American Economic Review, Vol. 67, No 4, pp. 639-652

Thaurel-Richard Michèle (1995), “Les progrès des élèves au CE2 en milieu rural”, Éducation et Formations, Vol. 43, pp.117-123

Woessman Ludger, West Martin R (2002), “Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS”, IZA Discussion Paper No 485

Appendix A: Class effects estimated with mixed models

The mixed model is written:

$$Y = X\beta + Z\gamma + \varepsilon$$

where γ and ε are Gaussian:

$$E\begin{bmatrix}\gamma \\ \varepsilon\end{bmatrix} = \begin{bmatrix}0 \\ 0\end{bmatrix} \text{ and } Var\begin{bmatrix}\gamma \\ \varepsilon\end{bmatrix} = \begin{bmatrix}G & 0 \\ 0 & R\end{bmatrix}$$

The endogenous variable Y is explained with covariates X multiplied by fixed-effects parameters β and with covariates Z multiplied by random-effects parameters γ.

For estimating class effects, the matrix Z is composed of class dummies. The vector γ is then a vector of random class effects. We assume that variance matrices G and R are diagonal: $G = \sigma_1^2 I$ and $R = \sigma^2 I$. G diagonal means that the random effects are uncorrelated.

G and R can be estimated by the method of restricted/residual maximum likelihood (REML). Coefficients are then determined with Henderson's mixed model equations:

$$\begin{bmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + \hat{G}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\beta} \\ \hat{\gamma}\end{bmatrix} = \begin{bmatrix}X'\hat{R}^{-1}Y \\ Z'\hat{R}^{-1}Y\end{bmatrix}$$

The coefficients are thus:

$$\hat{\beta} = \left[X'\left(\hat{R} + Z\hat{G}Z'\right)^{-1} X \right]^{-1} X'\left(\hat{R} + Z\hat{G}Z'\right)^{-1} Y$$

$$\hat{\gamma} = \left(Z'\hat{R}^{-1}Z + \hat{G}^{-1}\right)^{-1}\left\{ Z'\hat{R}^{-1} - Z'\hat{R}^{-1}X\left[X'\left(\hat{R} + Z\hat{G}Z'\right)^{-1} X \right]^{-1} X'\left(\hat{R} + Z\hat{G}Z'\right)^{-1}\right\} Y$$

Thus, if the eigenvalues of $\hat{G} \rightarrow \infty$, then $\hat{G}^{-1} \rightarrow 0$, and the system is identical to the one of the estimation of fixed effects.

Appendix B: Tables and figures

Table 1: Number of students and classes according to the type of the teacher

	Number of teachers	Number of pupils
Experienced teachers	96	2122
Trained novice teachers	65	1241
Untrained novice teachers	36	638
Total	197	4001

Table 2: Statistics on the student variables

Variable	N	Mean	Standard deviation
Month of birth	4001	6.52	3.32
One year younger than usual age	4001	0.02	0.13
Female	4001	0.47	0.50
Foreign nationality	4001	0.16	0.37
1 sibling	4001	0.37	0.48
2 siblings	4001	0.27	0.44
3 or more siblings	4001	0.27	0.45
Father's occupation: high	4001	0.32	0.47
Mother's occupation: high	4001	0.17	0.37
Pre-elementary school: less than 3 years	4001	0.22	0.42
Pre-elementary school: more than 3 years	4001	0.18	0.38
One repeated grade in pre-elementary school	4001	0.02	0.12
First grade repeated	4001	0.12	0.33
Second grade repeated	4001	0.07	0.25
Third grade repeated	4001	0.07	0.26
Initial test score in reading	3619	100.00	15.00
Initial test score in math	3683	100.00	15.00
Final test score in reading	3324	100.00	15.00
Final test score in math	3467	100.00	15.00

Note: High occupation corresponds to self-employed worker, executive, teacher, professor, technician, and foreman.

Table 3: Statistics on the teacher and class variables

Variable	N	Mean	Standard deviation
Experienced teacher	197	0.49	0.50
Trained novice teacher	197	0.33	0.47
Untrained novice teacher	197	0.18	0.39
Field of specialization: sciences	197	0.12	0.32
Field of specialization: unknown	197	0.40	0.49
Class size	189	23.92	4.06
Combination class	196	0.32	0.47
Rural	197	0.15	0.36
Semi-rural	197	0.18	0.39
Priority educational area (ZEP)	197	0.25	0.43

Note: The field of specialization is unknown when the teacher did not answer the question or when the teacher did not go to the university.

Table 4: Statistics on the student variables, for those students whose characteristics are all known

Variable	N	Mean	Standard deviation
Month of birth	2791	6.56	3.36
One year younger than usual age	2791	0.02	0.14
Female	2791	0.48	0.50
Foreign nationality	2791	0.16	0.37
1 sibling	2791	0.36	0.48
2 siblings	2791	0.28	0.45
3 or more siblings	2791	0.26	0.44
Father's occupation: high	2791	0.33	0.47
Mother's occupation: high	2791	0.17	0.38
Pre-elementary school: less than 3 years	2791	0.20	0.40
Pre-elementary school: more than 3 years	2791	0.16	0.37
One repeated grade in pre-elementary school	2791	0.01	0.11
First grade repeated	2791	0.11	0.31
Second grade repeated	2791	0.07	0.25
Third grade repeated	2791	0.06	0.24
Initial test score in reading	2791	100.42	14.77
Initial test score in math	2791	100.62	14.88
Final test score in reading	2791	100.51	14.86
Final test score in math	2791	100.51	14.90

Note: High occupation corresponds to self-employed worker, executive, teacher, professor, technician, and foreman.

Table 5: Statistics on the teacher and class variables, for those students whose characteristics are all known

Variable	N	Mean	Standard deviation
Experienced teacher	177	0.49	0.50
Trained novice teacher	177	0.34	0.48
Untrained novice teacher	177	0.17	0.38
Field of specialization: sciences	177	0.11	0.32
Field of specialization: unknown	177	0.41	0.49
Class size	177	23.93	4.07
Combination class	177	0.28	0.45
Rural	177	0.15	0.36
Semi-rural	177	0.18	0.38
Priority educational area (ZEP)	177	0.26	0.44

Note: The field of specialization is unknown when the teacher did not answer the question or when the teacher did not go to the university.

Table 6: Statistics on the student variables by type of teacher

	Experienced teachers		Trained novice teachers		Untrained novice teachers	
Variable	N	Mean	N	Mean	N	Mean
Month of birth	2122	6.52	1241	6.51	638	6.55
		(3.31)		(3.36)		(3.27)
One year younger than usual age	2122	0.02	1241	0.01	638	0.02
		(0.13)		(0.11)		(0.13)
Female	2122	0.47	1241	0.47	638	0.47
		(0.50)		(0.50)		(0.50)
Foreign nationality	2122	0.14	1241	0.20	638	0.18
		(0.34)		(0.40)		(0.39)
1 sibling	2122	0.40	1241	0.31	638	0.36
		(0.49)		(0.46)		(0.48)
2 siblings	2122	0.25	1241	0.28	638	0.28
		(0.44)		(0.45)		(0.45)
3 or more siblings	2122	0.24	1241	0.34	638	0.26
		(0.43)		(0.47)		(0.44)
Father's occupation: high	2122	0.35	1241	0.29	638	0.29
		(0.48)		(0.45)		(0.45)
Mother's occupation: high	2122	0.19	1241	0.14	638	0.13
		(0.39)		(0.35)		(0.34)
Pre-elementary school: less than 3 years	2122	0.21	1241	0.23	638	0.25
		(0.41)		(0.42)		(0.43)
Pre-elementary school: more than 3 years	2122	0.18	1241	0.16	638	0.18
		(0.39)		(0.37)		(0.38)
One repeated grade in pre-elementary school	2122	0.01	1241	0.02	638	0.01
		(0.12)		(0.13)		(0.12)
First grade repeated	2122	0.10	1241	0.15	638	0.14
		(0.30)		(0.36)		(0.35)
Second grade repeated	2122	0.07	1241	0.07	638	0.06
		(0.25)		(0.26)		(0.24)
Third grade repeated	2122	0.07	1241	0.07	638	0.06
		(0.26)		(0.26)		(0.24)
Initial test score in reading	1951	102.38	1109	97.03	559	97.61
		(14.40)		(15.26)		(15.12)
Initial test score in math	1990	101.55	1101	98.46	592	97.65
		(15.45)		(14.08)		(14.54)
Final test score in reading	1803	102.75	1050	97.34	471	95.41
		(14.32)		(15.17)		(15.00)
Final test score in math	1876	102.24	1057	98.32	534	95.45
		(14.42)		(15.09)		(15.38)

Note: Standard deviations are reported in parentheses. High occupation corresponds to self-employed worker, executive, teacher, professor, technician, and foreman.

Table 7: Statistics on the teacher and class variables by type of teacher

	Experienced teachers		Trained novice teachers		Untrained novice teachers	
Variable	N	Mean	N	Mean	N	Mean
Field of specialization: sciences	96	0.09	65	0.15	36	0.11
		(0.29)		(0.36)		(0.32)
Field of specialization: unknown	96	0.67	65	0.12	36	0.17
		(0.47)		(0.33)		(0.38)
Class size	88	24.85	65	22.88	36	23.53
		(3.60)		(4.07)		(4.65)
Combination class	95	0.23	65	0.35	36	0.47
		(0.42)		(0.48)		(0.51)
Rural	96	0.13	65	0.11	36	0.31
		(0.33)		(0.31)		(0.47)
Semi-rural	96	0.15	65	0.18	36	0.28
		(0.35)		(0.39)		(0.45)
Priority educational area (ZEP)	96	0.19	65	0.34	36	0.25
		(0.39)		(0.48)		(0.44)

Note: Standard deviations are reported in parentheses. The field of specialization is unknown when the teacher did not answer the question or when the teacher did not go to the university.

Table 8: Regression estimates of initial test scores, with all pupils (main results)

Dependent variable: initial scores in:	reading	reading	reading	math	math	math
Experienced teacher	4.70 **		4.12 **	3.26 *		3.31 *
	(1.75)		(1.44)	(1.91)		(1.78)
Trained novice teacher	-1.09		0.69	-0.12		1.62
	(1.86)		(1.50)	(2.04)		(1.84)
Class size		0.37 **	0.24 *		0.24	0.22
		(0.17)	(0.13)		(0.18)	(0.16)
Other variables	No	No	Yes	No	No	Yes
Class effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of students	3619	3550	3550	3683	3615	3615
Number of classes	195	187	187	195	187	187
Class level residual variance	65.58	70.14	37.37	83.33	82.16	62.47
Individual level residual variance	157.65	156.89	126.00	142.75	142.73	124.87

Note: Regression of the **initial** test scores of **all students**. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Table 9: Regression estimates of initial test scores, on the pupils with novice teachers (main results)

Dependent variable: initial scores in:	reading	reading	reading	math	math	math
Trained novice teacher	-1.10		0.81	-0.06		1.98
	(1.98)		(1.72)	(1.77)		(1.53)
Class size		0.22	0.25		0.08	0.18
		(0.23)	(0.20)		(0.20)	(0.18)
Other variables	No	No	Yes	No	No	Yes
Class effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of students	1668	1668	1668	1693	1693	1693
Number of classes	99	99	99	99	99	99
Class level residual variance	75.03	74.64	47.71	60.12	60.08	36.86
Individual level residual variance	162.74	162.72	139.84	146.46	146.45	134.68

Note: Regression of the **initial** test scores of **the students with novice teachers**. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Table 10: Estimated effects of teacher and class characteristics on final test scores, on the pupils with novice teachers (main results)

Dependent variable: initial scores in:	reading	reading	reading	math	math	math
Trained novice teacher	0.35		1.31	2.24		3.60 **
	(1.87)		(1.17)	(1.81)		(1.37)
Class size		-0.36 *	-0.34 **		-0.38 *	-0.34 **
		(0.21)	(0.13)		0.21	(0.15)
Other variables	No	No	Yes	No	No	Yes
Class effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of students	1521	1521	1354	1591	1591	1464
Number of classes	98	98	95	98	98	97
Class level residual variance	59.73	57.61	18.84	59.27	58.19	27.63
Individual level residual variance	178.19	178.16	56.53	178.13	178.09	69.73

Note: Regression of the **final** test scores of **the students with novice teachers**; the covariates include student and class characteristics (including initial test scores), together with teacher characteristics. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Table 11: Regression estimates of initial test scores, with all pupils (all results)

Dependent variable: initial scores in	reading		math	
Constant	96.42 **	(3.79)	96.48 **	(4.65)
Pupil characteristics:				
Month of birth	-0.36 **	(0.06)	-0.40 **	(0.06)
One year younger than usual age	4.25 **	(1.53)	2.23	(1.53)
Female	3.12 **	(0.39)	-0.76 **	(0.38)
Foreign nationality	-3.96 **	(0.62)	-2.35 **	(0.61)
1 sibling	-1.15	(0.71)	-0.95	(0.71)
2 siblings	-1.78 **	(0.74)	-1.12	(0.73)
3 or more siblings	-4.38 **	(0.79)	-3.15 **	(0.78)
Father's occupation: high	2.99 **	(0.47)	3.72 **	(0.47)
Mother's occupation: high	2.93 **	(0.57)	3.12 **	(0.57)
Pre-elementary school: less than 3 years	-2.04 **	(0.58)	-1.59 **	(0.58)
Pre-elementary school: more than 3 years	-0.20	(0.64)	0.76	(0.64)
One repeated grade in pre-elementary school	-3.08 *	(1.69)	-5.25 **	(1.59)
First grade repeated	-7.64 **	(0.63)	-4.65 **	(0.61)
Second grade repeated	-4.14 **	(0.79)	-1.30 *	(0.77)
Third grade repeated	-6.50 **	(0.79)	-4.90 **	(0.78)
Class characteristics:			3.31 *	(1.78)
Experienced teacher	4.12 **	(1.44)		
Trained novice teacher	0.69	(1.50)	1.62	(1.84)
Class size	0.24 *	(0.13)	0.22	(0.16)
Combination class	1.40	(1.28)	0.72	(1.57)
Rural	0.58	(1.69)	4.06 *	(2.08)
Semi-rural	0.97	(1.39)	1.08	(1.71)
Priority educational area (ZEP)	-4.63 **	(1.25)	-2.21	(1.55)
Number of students	3550		3615	
Number of classes	187		187	
Class level residual variance	37.37		62.47	
Individual level residual variance	126.00		124.87	

Note: The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Table 12: Regression estimates of initial test scores, on the pupils with novice teachers (all results)

Dependent variable: initial scores in	reading			math		
Constant	95.19	**	(5.62)	96.39	**	(5.08)
Pupil characteristics:						
Month of birth	-0.35	**	(0.09)	-0.45	**	(0.09)
One year younger than usual age	2.32		(2.53)	1.90		(2.53)
Female	3.26	**	(0.59)	-0.79		(0.58)
Foreign nationality	-3.30	**	(0.91)	-1.57	*	(0.87)
1 sibling	-1.47		(1.12)	-1.51		(1.10)
2 siblings	-2.15	*	(1.15)	-1.48		(1.13)
3 or more siblings	-4.58	**	(1.21)	-3.65	**	(1.18)
Father's occupation: high	2.73	**	(0.75)	3.72	**	(0.73)
Mother's occupation: high	2.82	**	(0.93)	3.44	**	(0.92)
Pre-elementary school: less than 3 years	-0.78		(0.88)	-0.57		(0.85)
Pre-elementary school: more than 3 years	-0.18		(1.08)	0.98		(1.02)
One repeated grade in pre-elementary school	-0.68		(2.55)	-0.94		(2.40)
First grade repeated	-5.81	**	(0.88)	-2.59	**	-0.8498
Second grade repeated	-3.96	**	(1.22)	-0.81		(1.16)
Third grade repeated	-6.42	**	(1.29)	-2.84	**	(1.24)
Teacher characteristics:						
Trained novice teacher	0.81		(1.72)	1.98		(1.53)
Field of specialization: sciences	3.37		(2.27)	3.41		(2.11)
Field of specialization: unknown	2.37		(2.31)	4.19	**	(2.07)
Class and school characteristics:						
Class size	0.25		(0.20)	0.18		(0.18)
Combination class	-0.95		(2.01)	-1.35		(1.80)
Rural	2.37		(2.91)	5.44	**	(2.61)
Semi-rural	1.51		(2.20)	2.44		(1.91)
Priority educational area (ZEP)	-4.82	**	(1.89)	-2.64		(1.67)
Number of students	1668			1693		
Number of classes	99			99		
Class level residual variance	47.71			36.86		
Individual level residual variance	139.84			134.68		

Note: Regression of the **initial** test scores of **the students with novice teachers**. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Table 13: Estimated effects of teacher and class characteristics on final test scores, on the pupils with novice teachers (all results)

Dependent variable: final scores in	reading			math		
Constant	51.07	**	(7.56)	59.53	**	(10.20)
Pupil characteristics:						
Month of birth	0.10		(0.06)	0.05		(0.07)
One year younger than usual age	4.30	**	(1.69)	3.66	**	(1.88)
Female	2.25	**	(0.43)	-0.83	*	(0.45)
Foreign nationality	0.36		(0.64)	1.16	*	(0.69)
1 sibling	-0.67		(0.78)	-0.74		(0.85)
2 siblings	0.11		(0.81)	-0.31		(0.87)
3 or more siblings	-0.10		(0.85)	0.22		(0.92)
Father's occupation: high	1.47	**	(0.52)	0.14		(0.57)
Mother's occupation: high	0.80		(0.65)	1.20	*	(0.72)
Pre-elementary school: less than 3 years	-0.40		(0.62)	-0.52		(0.67)
Pre-elementary school: more than 3 years	1.36	*	(0.78)	0.06		(0.81)
One repeated grade in pre-elementary school	-6.45	**	(1.93)	-3.88	**	(1.93)
First grade repeated	-5.39	**	(0.63)	-3.82	**	(0.67)
Second grade repeated	-3.83	**	(0.90)	-2.38	**	(0.93)
Third grade repeated	-2.32	**	(0.98)	-3.17	**	(0.98)
Initial test score	0.76	**	(0.02)	0.79	**	(0.02)
Teacher characteristics:						
Trained novice teacher	1.31		(1.17)	3.60	**	(1.37)
Field of specialization: sciences	0.21		(1.56)	3.48	*	(1.86)
Field of specialization: unknown	4.90	**	(1.50)	3.43	*	(1.80)
Class and school characteristics:						
Class size	-0.34	**	(0.13)	-0.34	**	(0.15)
Combination class	1.09		(1.37)	1.05		(1.58)
Mean of initial scores in the class	-0.21	**	(0.06)	-0.25	**	(0.08)
Standard deviation of initial scores in the class	-0.16		(0.18)	-0.67	**	(0.24)
Rural	-2.37		(1.94)	-4.76	**	(2.30)
Semi-rural	-1.67		(1.50)	0.44		(1.71)
Priority educational area (ZEP)	-1.74		(1.34)	-3.24	**	(1.48)
Number of students	1354			1464		
Number of classes	95			97		
Class level residual variance	18.84			27.63		
Individual level residual variance	56.53			69.73		

Note: Regression of the **final** test scores of **the students with novice teachers**; the covariates include student and class characteristics (including initial test scores), together with teacher characteristics. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Table 14: Heterogeneous effects of training

Quartiles within the classes of students' initial reading scores

Dependent variable: final scores in	reading	math
teachers' training	1.42	1.90
and low achieving students	(1.30)	(1.45)
teachers' training	1.19	4.37 **
and medium low achieving students	(1.25)	(1.45)
teachers' training	1.00	4.05 **
and medium high achieving students	(1.25)	(1.45)
teachers' training	1.67	4.73 **
and high achieving students	(1.32)	(1.49)
Other variables	Yes	Yes
Class effects	Yes	Yes
Test de Fisher Q2 = Q1 (p value)	0.75	0.00
Test de Fisher Q3 = Q1 (p value)	0.63	0.01
Test de Fisher Q4 = Q1 (p value)	0.82	0.00
Number of students	1354	1403
Number of classes	95	96
Class level residual variance	18.85	27.15
Individual level residual variance	56.62	69.12

Note: Regression of the **final** test scores of **the students with novice teachers**; the covariates include student, class characteristics and dummy variable of trained teachers, broken down according to the quartiles. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Quartiles of class means of initial reading scores

Dependent variable: final scores in	reading	math
teachers' training	-1.25	0.68
and low achieving classes	(1.88)	(2.03)
teachers' training	0.50	5.70 **
and medium low achieving classes	(1.67)	(1.90)
teachers' training	0.79	3.69 *
and medium high achieving classes	(1.70)	(2.00)
teachers' training	4.26 **	4.62 **
and high achieving classes	(1.76)	(1.98)
Other variables	Yes	Yes
Class effects	Yes	Yes
Test de Fisher Q2 = Q1 (p value)	0.37	0.02
Test de Fisher Q3 = Q1 (p value)	0.36	0.21
Test de Fisher Q4 = Q1 (p value)	0.03	0.13
Number of students	1354	1403
Number of classes	95	96
Class level residual variance	18.62	26.16
Individual level residual variance	56.47	69.59

Note: Regression of the **final** test scores of **the students with novice teachers**; the covariates include student, class characteristics and dummy variable of trained teachers, broken down according to the quartiles. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Table 15: Heterogeneous effects of class size

Quartiles within the classes of students' initial reading scores

Dependent variable: final scores in	reading	math
class size	-0.29 **	-0.43 **
and low achieving students	(0.13)	(0.15)
class size	-0.33 **	-0.34 **
and medium low achieving students	(0.13)	(0.15)
class size	-0.35 **	-0.32 **
and medium high achieving students	(0.13)	(0.15)
class size	-0.36 **	-0.28 *
and high achieving students	(0.13)	(0.15)
Other variables	Yes	Yes
Class effects	Yes	Yes
Test de Fisher Q2 = Q1 (p value)	0.17	0.00
Test de Fisher Q3 = Q1 (p value)	0.12	0.00
Test de Fisher Q4 = Q1 (p value)	0.17	0.00
Number of students	1354	1403
Number of classes	95	96
Class level residual variance	18.65	27.35
Individual level residual variance	56.59	68.51

Note: Regression of the **final** test scores of **the students with novice teachers**; the covariates include student, class characteristics and class size variable, broken down according to the quartiles. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. $**p<0.05$ $*p<0.10$

Quartiles of class means of initial reading scores

Dependent variable: final scores in	reading	math
class size	-0.41 **	-0.54 **
and low achieving classes	(0.16)	(0.17)
class size	-0.36 **	-0.35 **
and medium low achieving classes	(0.13)	(0.15)
class size	-0.40 **	-0.35 **
and medium high achieving classes	(0.14)	(0.15)
class size	-0.23	-0.22
and high achieving classes	(0.15)	(0.16)
Other variables	Yes	Yes
Class effects	Yes	Yes
Test de Fisher Q2 = Q1 (p value)	0.59	0.02
Test de Fisher Q3 = Q1 (p value)	0.97	0.05
Test de Fisher Q4 = Q1 (p value)	0.25	0.01
Number of students	1354	1403
Number of classes	95	96
Class level residual variance	18.61	25.31
Individual level residual variance	56.47	69.60

Note: Regression of the **final** test scores of **the students with novice teachers**; the covariates include student, class characteristics and class size variable, broken down according to the quartiles. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Table 16: Effect of class size in the ZEP schools

Dependent variable: final scores in	reading	math
class size	-0.22	-0.12
	(0.14)	(0.16)
class size in the ZEP schools	-0.61 *	-1.09 **
	(0.32)	(0.35)
Number of students	1354	1464
Number of classes	95	97
Class level residual variance	18.41	24.69
Individual level residual variance	56.49	69.65

Note: Regression of the **final** test scores of **the students with novice teachers**; the covariates include student, class characteristics and class size variable, broken down according to ZEP schools and other schools. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Table 17: Effect of class size estimated with or without instrumental variable

	All classes			
	OLS	IV	IV (29-40)	IV (24-45)
reading	-0.36**	-0.39**	-0.18	-0.44**
	(0.10)	*(0.10)*	*(0.49)*	*(0.22)*
math	-0.43**	-0.44**	-0.84*	-0.67**
	(0.12)	*(0.12)*	*(0.45)*	*(0.23)*

	Without combination classes			
	OLS	IV	IV (29-40)	IV (24-45)
reading	-0.35**	-0.36**	-0.24	-0.32
	(0.12)	*(0.12)*	*(0.33)*	*0.20*
math	-0.49**	-0.50**	-0.80*	-0.54**
	(0.14)	*(0.14)*	*(0.43)*	*(0.21)*

Note: Regression of the **final** test scores of **all the students**; the covariates include student, class characteristics and instrumented variable of class size. The two last columns correspond to the estimation on the sub-samples restricted to some sizes of the enrollment. The coefficients are estimated through a mixed model, with class effects. Standard errors are reported in parentheses. **p<0.05 *p<0.10

Figure 1: Evolution of the number of individuals directly selected for entry into the teacher training college and of the number of individuals recruited on the waiting list, between the years 1986 to 1992

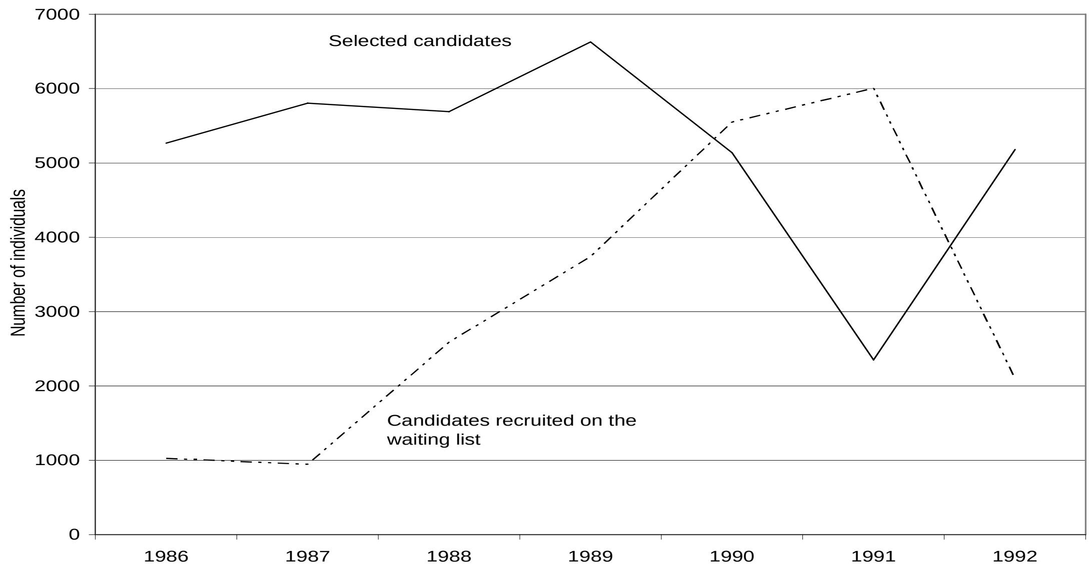

Figure 2: Class sizes and mean scores in reading for classes with experienced teachers

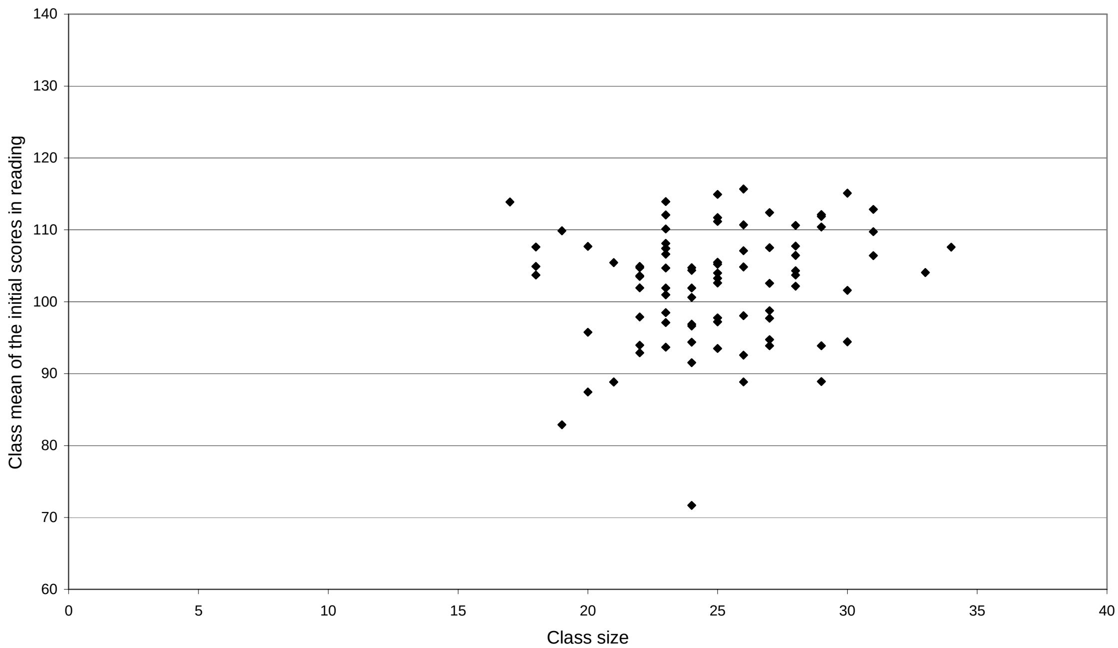

Figure 3: Class sizes and mean scores in reading for classes with novice teachers

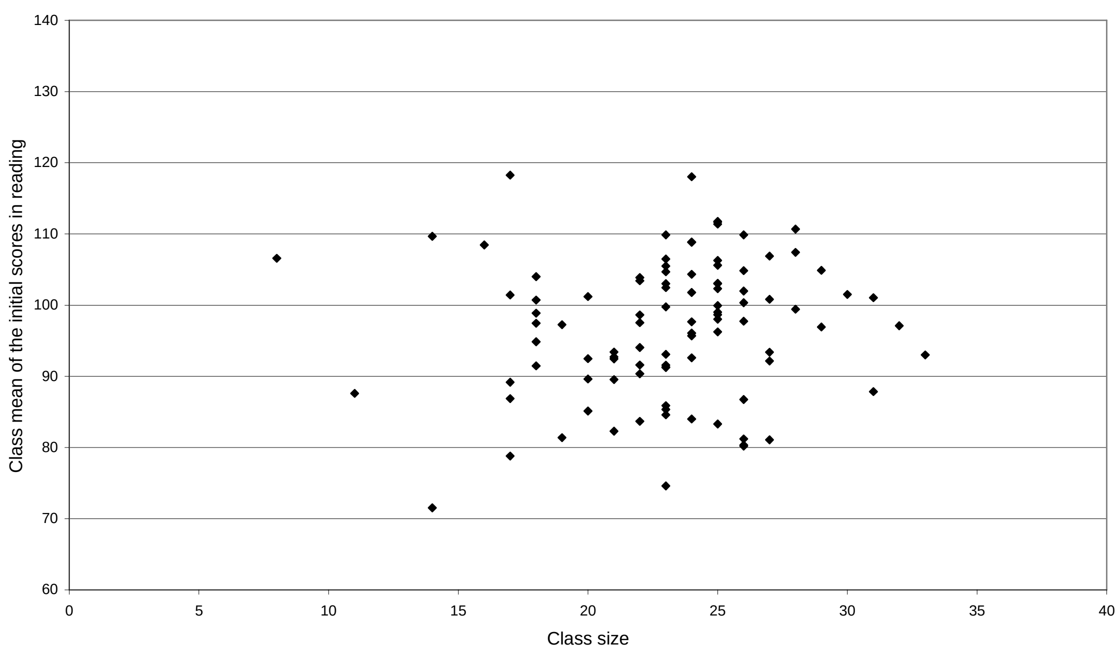

Figure 4: Enrollment and number of third-graders per class

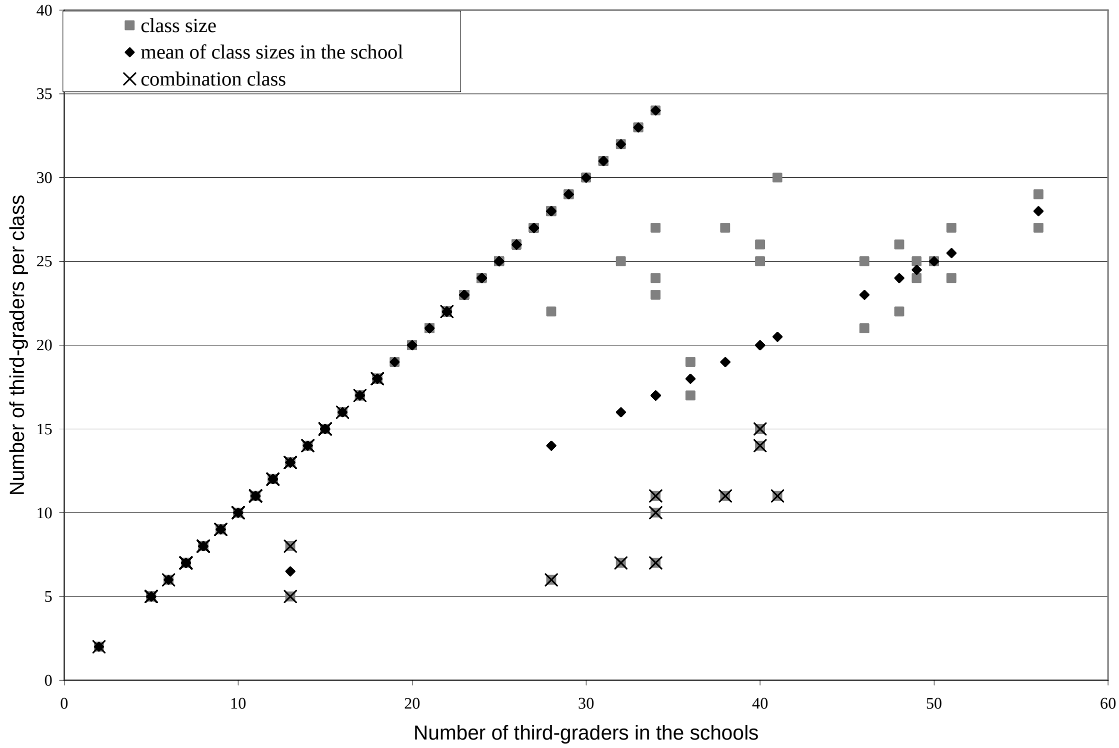

Figure 5: Enrollment and size of third grade classes, excluding combination classes

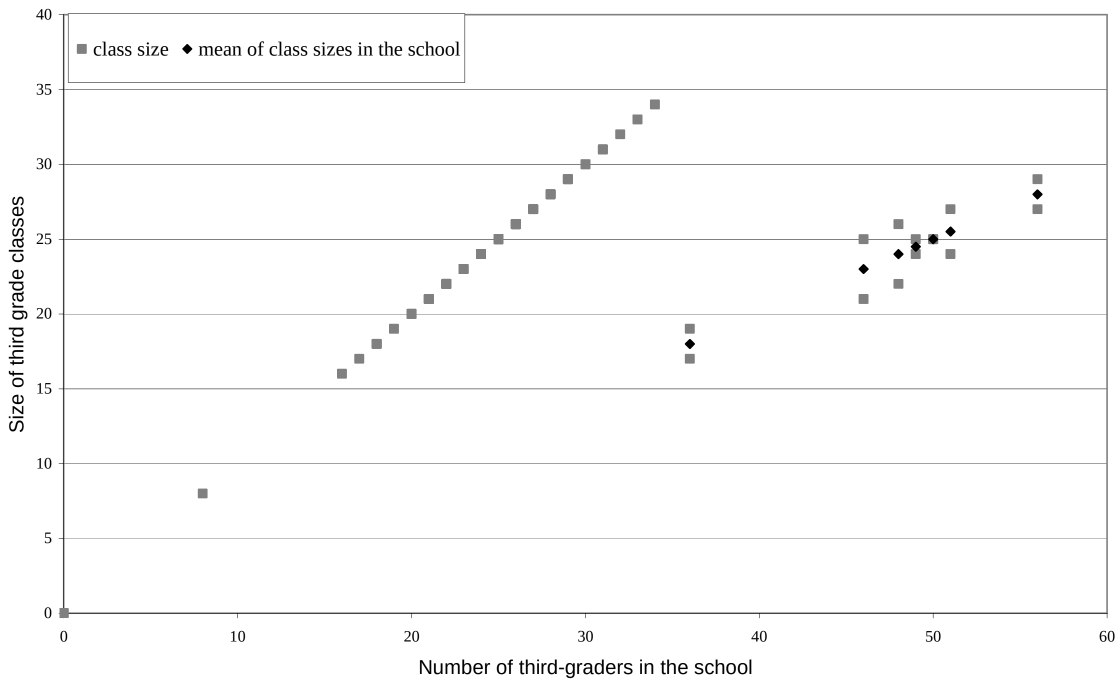

Figure 6: Enrollment of third-graders and students who are combined in a class with third-graders and class size