Estimating class-size effects using variation in subject-specific classes

Eskil Heinesen
AKF, Danish Institute of Governmental Research
Nyropsgade 37, DK-1602 Copenhagen V, Denmark
e-mail: esh@akf.dk

May 2007

Abstract

When considering alternative subjects among which students have to choose, within-school variation over time in the size of subject-specific classes reflects random variation in the share of students preferring one subject over another. At 7th grade in Danish schools, students must choose between German and French. The effect of French class size on French examination marks is estimated, and highly significant negative effects are found which are larger for academically weak students and for boys. Since all other subjects are taught in other (basic) classes, the marks attained in those subjects can be used to test and control for possible selection effects.

Key words: School quality; class size; human capital investment; fixed effects; panel data

JEL codes: C23, I2, I21

Martin Browning, Helena Skyt Nielsen, Beatrice Scindler Rangvid, Michael Svarer and participants in the Workshop at Centre of Applied Microeconometrics, University of Copenhagen, 14 December 2006, are gratefully acknowledged for helpful comments. The Danish Social Science Research Council is acknowledged with gratitude for support.

1. Introduction

Class size is generally considered to be a central aspect of school quality. It is of great concern to parents and is directly affected by political choices. To what extent class size matters for student education outcomes is an issue of great importance in public choice of school quality; there is a large literature on the estimation of causal effects of class size on test scores and other student outcomes (see, e.g., Hanushek, 1996; Card and Krueger, 1996; Hedges and Greenwald, 1996; Dearden et al., 2002; Dustmann et al., 2003; and Krueger, 2003).

Estimation of causal effects of class size or other school inputs is difficult because of endogeneity problems. Whereas existing class-size literature focuses on variation in the size of basic classes in which most subjects are taught, the present paper focuses on variation in the size of subject-specific classes involving alternative subjects between which students choose, since this variation should be less affected by endogeneity problems.

One important source of endogeneity is the selection response of parents to poor school quality, for reasons such as large class size, low quality of teachers, or bad social relations between students in the class. Thus, the size of basic classes may be affected by the quality of teaching or the quality of social relations between students in the class. For instance, parents who feel strongly about their children's education and well-being in school may move their children to other schools if they are not satisfied with school quality. Selection response of parents to a large class size or other aspects of school quality may also be a problem in experimental studies; see the discussion in Krueger (1999) and Hanushek (1999).[1] The same is true for quasi-experimental studies, such as those using maximum class-size rules in a regression discontinuity framework (see, e.g., Angrist and Lavy, 1999). This problem is particularly important when analysing class-size

[1] Analysing attrition from the treatment and control groups of the STAR experiment, Krueger (1999) concludes that selection response to a large class size is not an important problem for that particular experiment. However, Hanushek (1999) argues that non-random attrition and other implementation issues may bias estimated class-size effects based on this experiment.

effects at higher grades, since parents will then have had several years to move their child to another school.

The trend toward freer school choice in many countries in recent years has increased this endogeneity problem. Thus, if it is easy for parents to respond to a large class size (or other aspects of school quality) by moving their child to another school where class size at that grade is smaller (or other aspects of school quality are better), the question of which students are at schools with small and large class sizes may not be determined randomly. Most empirical methods for estimating class-size effects are less credible if it is easy for parents to get their children into public schools outside the catchment area of their residence, or if private schools are an option considered by many parents (e.g. because of low costs to parents due to public subsidies to private schools).

Another important source of endogeneity stems from choices of school administrators. Thus, schools receiving students from other schools may place the new students in classes which function well even if there are parallel classes which are smaller. Also, less able students may often be placed in smaller classes; see West and Wössmann (2006).

Class size in basic classes may therefore be strongly affected by choices of parents and school administrators, and smaller classes may often be small because they are of low quality or have many disadvantaged students. These endogeneity problems are not likely to be important to class size in subject-specific classes, at least not for schools where there is only one class in the given subject in a given year and grade. Thus, parents would probably not consider moving their child to another school merely because they are dissatisfied with the quality of teaching in a single subject (for instance, because they consider the subject-specific class to be too large). Also, since the allocation of students in subject-specific classes is determined by students' preferences for different optional subjects, class size in subject-specific classes is not determined by decisions of school administrators, at least where there is only one subject-specific class for each school/year cell.

Thus, analysis of the effects of class size in subject-specific classes, for instance on test scores or examination marks in the given subject, makes it possible to avoid endogeneity problems, including selection response to larger classes, which is an important issue in most studies of class-size effects, including experimental and quasi-experimental studies. A further advantage of this identification strategy is that test scores or marks in other subjects, which are taught in other (basic) classes, may be used to test and control for possible selection effects.

A further issue in the political economy of education is who benefits from increases in school resources, such as reductions in class size. In the Scandinavian welfare states and in many other countries, reducing inequalities in society is a leading political priority, and a high quality of public schools is often seen as an important means to reduce inequality in education (which in turn is essential for reducing economic inequality later in life) and promote intergenerational mobility. Therefore, it is important to investigate whether increasing school resources have larger effects for academically weak students or students from disadvantaged backgrounds. By analysing the effect of class size on test scores or marks in subjects taught in subject-specific classes, it is possible to study interactions between class-size effects and student academic ability (measured by test scores or marks in other subjects) and other student characteristics such as parental background and gender.

The present work investigates class-size effects in subject-specific classes, namely French classes, using Danish administrative register data for three cohorts of students. These micro data contain information on examination marks at the end of 9th grade, enrolment, school characteristics and parental background. In Denmark, all students in primary and lower secondary school (grades 1-9) study two foreign languages. They learn English from 4th grade onwards, and German or French from 7th grade onwards. All schools offer German, and about 24% of students attend schools offering both German and French from 7th grade. Within this category of schools, the size of German and French classes at a given school in a given year depends on how many students choose one or the other. At schools with two or more parallel basic classes, French is taught in French classes with students from several basic classes, and similarly for German. All subjects other than German and French are typically taught in basic classes.[2]

It is not possible to predict before grade 7 (or the end of grade 6) how many students will choose German or French. It is reasonable to suppose that students who choose French (for example) will not change school due to large class size or low teacher or peer group quality in the French class specifically, since the quality of basic classes is far more important. By using variation in subject-specific classes we can therefore avoid important selection problems. Also, we identify class-size effects from variation over time within schools only, thereby taking account of time-constant unobserved differences between schools and catchment-area populations. This may be important due to between-school variation in unobserved dimensions of student ability and school quality which may be correlated with the numbers choosing French at different schools. For instance, if French teachers are (known to be) very good at some schools, many students may choose French at these schools leading to large class sizes in French, and examination marks may be high in spite of large class sizes.

Reducing class size is found to have highly significant and substantial positive effects on examination marks, also when controlling for marks in other subjects. The effect is significantly larger for academically weak students (where academic ability is measured by marks in other subjects) and for boys. The effect is also larger for students whose parents do not have an academic education, although the class-size effect does not depend on parental education when we control for marks in other subjects.

[2] Due to data issues this paper focuses on class-size effects in French, and not in German; this is explained in Section 3.

Section 2 describes the institutional setting and Section 3 the data. Section 4 discusses the econometric methods used, and Section 5 presents the results. Section 6 contains comparisons with other estimates of class-size effects in the literature, and Section 7 gives conclusions.

2. Institutional setting

In Denmark, compulsory education runs from 1st to 9th grade (age 7-15 years). Public schools are free. Expenditure on public schools is (like most other municipal expenditure) financed by municipal taxes, primarily income tax, but extensive grants and equalization schemes eliminate the greater part of financial inequalities between municipalities. Private schools (including boarding schools) are heavily subsidized by the state and the municipalities; parents pay on average a tuition fee of about 15% of the costs.

At 1st grade almost 90% are enrolled in public schools, and the remaining 10% attend private schools. However, a few per cent of those starting in public schools change to private schools during their school career, and in 9th grade some students change to boarding schools. For the three cohorts analysed in this paper, 76% attended public schools in 9th grade, 11% attended boarding schools, and 13% attended other private schools.

At 7th grade students start learning their second foreign language: German or French (the first is English, which is taught from 4th grade onwards). The number of weekly lessons in German and French is 3 in 7th grade and 4 in 8th and 9th grade. All schools offer German, and about 24% of students attend schools which offer both German and French from 7th grade; of this group of students, about 28% choose French.

At the end of compulsory schooling (9th grade) there is an examination in Danish, mathematics, science (physics and chemistry), English, and German or French. There are three written exams (two in Danish and one in math) and four oral exams (in Danish, math, science, and

French or German). Oral examinations are conducted by the teacher and an external moderator. Students also receive marks for their year's work, which are given by the teacher. Only students choosing academic upper secondary school after 9th grade actually need to take the examination, and some students who do not take the examination in the year they complete 9th grade may take an examination after an optional 10th grade at lower secondary school. About 98% out of those enrolled in 9th grade at the beginning of the school year receive marks for the year's work in Danish and math, and 99% of those take the examination in these subjects. Fewer take the examination in other subjects. Only about 90 and 86% of the students with marks for the year's work in German and French (respectively) take the examination in these subjects.

3. Data and descriptive statistics

Danish data are used comprising several administrative registers which are merged using personal and school registration numbers. Data are available for all students in 9th grade in the three school years 2001/2002 – 2003/2004. For each student who attended lessons in a given subject, the database holds marks for the year's work, and for those who have taken the examination at the end of 9th grade it includes their examination marks. There is also information on the school attended in 9th grade, and extensive information on personal characteristics (including gender and immigrant status), and family background (siblings, family structure, and parents' education, age, labour-market status, income, unemployment and housing conditions). All family background variables are measured in the calendar year the child reached 12 years of age.

The total number of students finishing 9th grade in 2002-2004 was 154,156. Excluding schools which do not offer French from 7th grade onwards there are 36,687 students.[3] Of this group,

[3] Among the 117,469 deleted observations are 24,629 students attending schools which do not offer French from 7th grade onwards, but as an additional voluntary subject in 8th and 9th grade only. In these schools all students have to attend German from 7th grade onwards, and they may choose to attend French from 8th grade as an additional voluntary subject. About 15% of students in these schools choose French, comprising a selective group of high-ability students.

10,314 studied French (i.e., they have marks for the year's work in French in 9th grade), and 8844 took the examination (i.e., they have also examination marks in French). There are no data on the number of French classes at each school. The analysis is therefore restricted to schools in which the number of students attending French in 9th grade is 20 or less. In that case one can be quite sure that only one French class is established so that class size is equal to the number of students attending French, i.e. the number of students having marks for the year's work in French.[4] The restriction that at most 20 students attend French reduces the sample to 6612 observations. Further, we only analyse school/year cells in which some students attend lessons in German, so that class size in French is not equal to enrolment. This reduces the final estimation sample to 6507 students. The size of French classes in this sample is only weakly correlated with grade enrolment; the within-school correlation coefficient is 0.36.[5]

The explanatory variable of primary interest in this paper is class size, which is measured for each school/year cell. There are 782 school/year cells in the estimation sample, and 375 schools. 114 schools are observed in one year only, 115 in two years, and 146 in three years. This gives 407 observations for identifying class-size effects from within-school variation $(115+2\times146)$.

The marking scale in the Danish educational system has values between 0 and 13, excluding 1, 2, 4, and 12. Mark 0 is given for 'completely unacceptable performance', and 13 for 'exceptionally independent and excellent performance'. To pass an exam, students must score at

[4] The cut-off at 20 is chosen from the relation between grade enrolment and the number of basic classes. Only 6% of schools with 20 9th grade students have two classes, whereas the same is true for 15% of schools with an enrolment count of 21, and for 18, 30 and 47% of schools with enrolment counts 22, 23 and 24, respectively.

[5] Similar sample restrictions result in an estimation sample for the analysis of class-size effects in German classes where the within-school correlation between German class size and grade enrolment is much higher (0.64), and the sample size is much smaller. In particular, the number of schools with observations for more than one year becomes very small (115). Also, variation in the size of German classes is smaller than the variation of French classes, and the vast majority of German students in the estimation sample are in classes close to the maximum of 20 (only 9% are in classes with less than 10 students, and 76% are in classes with more than 12 students; the corresponding percentages for French students are 20 and 52, respectively). Consequently, estimates of class-size effects in German classes are imprecise and very sensitive to changes of model specification (e.g. to control for grade enrolment). This is why we focus on class-size effects in French classes only.

least 6 – the three lowest marks (0, 3 and 5) are for ‘unsatisfactory performances’. Mark 8 is for ‘average performance’. The histogram in Figure 1 shows the distribution of examination marks in French for students in the estimation sample. Variation is quite large, although 52% scored 7, 8 or 9. The top marks of 11 or 13 are obtained by 9%, and 11% score 0, 3 or 5. The marks at the very top and bottom of the scale are seldom used: Only 1.5% score 13, 2.1% score 3, and only 6 students out of 6507 scored 0. Marks for the year’s work (not shown) vary considerably less: the variance is less than half that of examination marks, both between and within schools. Thus, within a class average marks for the year’s work are typically close to 8, and these marks may be less ‘objective’ than examination marks since they are given by one person only (the teacher). Only examination marks are analysed in this paper.

[Figure 1 about here]

Figure 2 shows the number of observations (students) by class size in French classes. There are few observations for class sizes below 9, particularly below 5, or above 18. Figure 3 shows average examination marks in French by class size in class-size intervals of two; also shown are average residuals from a school fixed-effects regression of marks on time dummies, and residuals from a similar regression which also includes a large set of controls (see below). The average marks increase when class size increases from 1-2 students to 5-6 students, but there is a negative trend when class size increases beyond this level. It is this negative trend which is important empirically, since there are very few observations for class size below five (see Figure 2). This negative trend is much more pronounced when examination marks are corrected for school and year fixed effects (and other controls) – see the other two graphs in Figure 3. The fact that these two graphs are very similar indicates that, once school (and year) fixed effects are controlled for, it is not very important for the estimation of class-size effects whether we also control for individual family background, grade enrolment, etc., even if these controls are themselves highly significant. The large differences between the curve for observed marks and the curves for marks corrected for school fixed effects indicate that the selection mechanisms associated with unobserved differences between schools may be very important, and that failing to take account of this selection is liable to bias estimates of class-size effects downwards.

[Figure 2 about here]

[Figure 3 about here]

The controls included in the estimation which gives rise to the average residuals in the last graph of Figure 3 are also used in the estimations discussed below. Controls at the individual level are: a gender dummy, immigrant/second generation immigrant dummies, and family background variables (32 variables for siblings, family structure, and parents' education, age, labour-market status, income, unemployment and housing conditions). Controls at school/year level are: grade enrolment, enrolment squared, and the fraction of students who do not attend either French or German lessons (which is an indicator of the fraction of academically less able students).[6] In addition, year dummies are included in all regressions (though they are clearly insignificant). Summary statistics for controls are shown in the appendix.

4. Empirical methods

The econometric analyses of this paper are based on the model:

[6] This last variable is not significant at the 5% level in any of the estimations; excluding it from the model does not change the results in any substantial way.

$$y_{ist} = X_{ist}\beta + n_{st}\alpha + \eta_t + \mu_s + \varepsilon_{ist}, \tag{1}$$

where y_{ist} is the French examination mark of student *i* of cohort *t* in school *s*, X_{ist} are controls (at individual and school/year level), n_{st} is French class size, η_t and μ_s respectively denote unobserved year and school fixed effects, and ε_{ist} is an error term.

Taking account of school fixed effects wipes out all time-constant unobserved differences between schools. Thus, class-size effects are identified by variation over time within the same school. Within-school variation in the size of French classes is unpredictable at the time of school choice, and is driven by random variation in the number of students preferring to learn French instead of German.

To interpret any estimate of α, which is the parameter of prime interest, as a causal effect of class size on examination marks, three identifying assumptions must hold: (a) students do not change school if they are dissatisfied with the quality of teaching in French (for instance, because the French class is large); (b) there is no correlation between the underlying ability to learn French and the size of French classes; and (c) it is not possible to predict the quality of teaching in French and German classes. The Introduction argued that assumption (a) of no selection response to a large class size (or other quality factors of French teaching) is likely to hold. Thus, even though students and parents may believe that class size is an important aspect of school quality and react to a large size of basic classes, they will probably not move to another school just because of a large class size in a single subject with 3-4 lessons per week.

Assumption (b) may be violated if the number choosing French is correlated with these students' general ability or their ambition regarding learning foreign languages. For instance, it might be that when very few students at a school in a given year choose French, this is typically a group of very able and ambitious students; whereas if many choose French, they are on average less able. Estimates of class-size effects may then be upward biased even if we use within-school variation only. However, we are able to test this by regressing marks in other subjects on class size in French (i.e., replacing marks in French with marks in other subjects on the LHS of equation (1)). If we find a (significant) correlation between class size in French and marks in other subjects, we are able to control for this problem by including marks in other subjects as additional regressors in (1) when estimating models for marks in French, or by estimating models for the difference between marks in French and marks in other subjects.

Assumption (c) would be violated if students and parents know which teachers will teach the French and German classes and if they know the relative quality of those teachers. If, for instance, a really good French teacher is expected to teach the French class, more students may choose French. This would bias the estimated effect of class size downwards. In many schools it may not, however, be possible to predict which teachers will teach a given cohort, and the relative quality of French and German teachers may be unknown. Moreover, in schools with only one French class at each of the three grades 7-9, it is likely to be the same French teacher who teaches all cohorts. When this is the case, the school-fixed effects take account of the problem. Since we only use school/year cells with rather few French students, and therefore (presumably) only one French class, this may be the case for many schools in the sample. (Of the 407 school/year cells used to identify class-size effects, 292 are from schools with a single French class, of at most 20 French students, in all three years in the data period; see section 3.) Nevertheless, it is not possible to rule out (or test) the possibility that class-size estimates may be downward-biased because assumption (c) may not always hold.

As stated in the introduction, the fact that the analysis is restricted to school/year cells with only one French class means that there are no selection problems due to school administrators' possible placement of less able students in smaller classes.

5. Estimation results

Table 1 shows the estimated class-size coefficients from within-school regressions based on equation (1). All regressions include year dummies. The six regressions in Table 1 differ according to whether other controls are included: enrolment (in 9th grade), enrolment squared, and individual controls (i.e., 35 variables for individual and family characteristics, and in addition the fraction of students at 9th grade in the school who do not have marks for the year's work in either French or German). Results are shown with and without control for enrolment (and its squared value), since enrolment and the size of French classes are positively correlated (the within-school correlation coefficient is 0.36), especially for school/year cells with enrolment below 30, and variation in enrolment (even within schools) might be correlated with unobserved student ability. The table shows that in all six regressions the estimated effect of class size in French classes on marks in French is highly significant. Point estimates are about -0.06,[7] with t-ratios of about 4. Thus, when school fixed effects are controlled for, it is not important for class-size estimates whether we also control for enrolment and individual background characteristics.[8] School fixed effects are highly significant. An F test that they are all zero is clearly rejected. For instance, in the first regression of Table 1: $F(374, 6091) = 2.27$; $P<0.0001$.

[7] The standard deviation of the examination marks is 1.97 in the distribution of marks of individual students, and 1.05 in the distribution of average marks in school/year cells. Thus, a class-size coefficient of -0.06, which is the expected reduction in marks when class size is increased by 1, is approximately equal to the effect measured in standard deviations in the distribution of average marks in school/year cells, whereas the 'effect size' in terms of the distribution of individual marks is about -0.03.

[8] There is a similar finding in Krueger (1999).

[Table 1 about here]

Enrolment is not significant in any of the six estimations. Individual background characteristics are highly significant. When they are included, R^2-within is 0.10, and when they are excluded it is 0.005.[9] The estimated parameters of all variables in the first regression of Table 1 are shown in the Appendix, together with their t-ratios. The most significant variables are for gender, family structure, and parental education. Females have 0.76 higher marks than males, and students from broken families have 0.24 lower marks than students from intact families. Students whose parents have some form of academic education, i.e. upper secondary school or further or higher education, obtain significantly higher marks than students whose parents have no education beyond compulsory. Furthermore, the education of the mother has larger and more significant effects than the education of the father.

The standard errors reported in Table 1 and in the following tables are robust, i.e., they are corrected for clustering within schools. Alternative estimations based on 'collapsing' the data to a dataset of means of school/year cells give similar results for class-size point estimates and their standard errors.

Tests of linearity of class-size effects

The model (1) and the corresponding estimations in Table 1 assume a linear effect of class size. However, there are several reasons why class-size effects might be non-linear. The 'disruption model' of Lazear (2001) predicts (numerically) decreasing effects. Also, class-size effects are estimated here over a range including very small classes, and the effect of reducing class size from 6 to 2 (for instance) may be different from the effect of reducing it from 20 to 16. Figure 3 indicates

[9] 'R^2-within' is calculated as R^2 from an OLS regression on the within-transformed variables.

that reducing class size below 5 might have negative effects (whereas reducing it in the range 20 to 5 may have positive effects), although the number of observations is rather low for very small class sizes (see Figure 2). One reason why very small classes might have negative effects is that students for social reasons prefer classes with at least a few other students, and that this factor affects learning. Accordingly, several tests were conducted for non-linearity. Linearity was first tested against a very general alternative by replacing the class-size variable, *n*, in equation (1) by 19 dummy variables, one for each value of class size between 1 and 19 (leaving 20 as the reference category): $d_j = 1$ if $n = j$, d_j=0 otherwise, $j = 1,...,19$. Denoting the dummy coefficients of d_j by α_j, linearity in this model is implied by the following 18 linear restrictions:

$$\alpha_j - \alpha_{j+1} - \alpha_{19} = 0, \quad j = 1,...,18 \tag{2}$$

Upon estimating this model and testing the 18 restrictions in (2) using an F test, linearity cannot be rejected. Thus, including individual controls (corresponding to the first three regressions of Table 1), the F(18, 374) test statistic is between 0.74 and 0.75, depending on whether we control for enrolment and enrolment squared, with P values between 0.77 and 0.75; if we exclude individual controls, the F statistic is 0.97, with P values of 0.49-0.50.

Next, to increase the power of the test, the number of dummy variables for class size was reduced. Again, linearity is not rejected; see the Appendix.

In a third test of linearity, the 'continuous' class-size variable *n* is put back into the model, but a separate effect of very small classes is also allowed by including an extra dummy variable for class sizes between 1 and 4. The results of estimating (1) with this extra dummy variable are shown in Table 2 for the six combinations of controls used in Table 1. In the first three regressions, including individual background controls, the small-class dummy is not significant (numerical t-ratios are between 1.2 and 1.3), whereas it is marginally significant in the last three equations without individual controls (the numerical t-ratios are between 2.0 and 2.2). The sign of the dummy coefficient is negative, as expected. The point estimates indicate a negative effect of class size less than 5 (compared to the linear class-size function) of value 0.2 in the first three regressions, and 0.4 in the last three regressions. The point estimates of the linear class-size effect in Table 2 are larger than in Table 1: the difference is about 6% (0.004) including individual controls, and 11% (0.007) excluding these controls.

[Table 2 about here]

When estimating spline functions in class size the results are similar. For instance, defining splines by the knots 5, 10 and 15, the four slope estimates (including all controls) are 0.4, -0.9, -0.5 and -0.5, but the differences are not significant. Upon using a specification corresponding to the models estimated in Table 2, i.e. using only one knot at 5, equality of slopes is not rejected when individual controls are included; it is rejected at the 5% level when these controls are not included, but only marginally.

Finally, including both n and n^2 in (1) the squared term is clearly insignificant. In the estimations that follow, linearity of class-size effects is assumed. For instance, the dummy for class size less than 5 is not included. Even if it is marginally significant in Table 2 when individual controls are not included, it is not significant when they are included. In regressions controlling for examination marks in other subjects, and in differences-in-marks regressions, it is clearly insignificant; also when individual controls are excluded. If it is included anyway, it always results in a marginal increase in the estimated linear effect of French class size on French marks (for class size in the range 5-20).

Test for selection effects

As discussed in Section 4, we can test whether class size in French is correlated with the academic ability of students by regressing marks in other subjects on class size in French. In such regressions, class size in French should have significant negative 'effects' if the highly significant class-size effects in Tables 1 and 2 are due to selection effects such that small French classes are associated with high ability of the students. Even in the absence of selection effects there may still be negative effects of French class size on marks in other subjects, due to positive spill-over effects from what is learned in French classes to skills in other subjects. For instance, the higher your skills in one language (e.g. French), the easier it is to gain skills in another language (e.g. English or Danish). More general spill-over effects or externalities from class size (or other school quality dimensions) in one subject to skills in other subjects might also be present (e.g. by affecting the extent to which students are inspired to do well also in other subjects). However, spill-over effects to other subjects are expected to be substantially smaller than the direct effect of French class size on French marks, and spill-over effects to completely different subjects such as math are expected to be very small.

Table 3 shows estimation results corresponding to Table 1, except that French marks are replaced by examination marks in other subjects: English, Danish (oral, written, and spelling), math (oral and written), science, and an average of all these marks. Some students in the estimation sample do not have examination marks in all subjects. This is why the number of observations in the estimations is not the same for different subjects. In particular, quite a few students do not take the science examination. Average marks (used in the first four estimations in the table) are for each individual calculated using all his or her non-missing marks. Four estimation results for each subject are shown. Estimation results controlling for both enrolment and enrolment squared are not shown, since the squared term is clearly insignificant in all estimations where it is included. The inclusion of enrolment squared gives coefficients of French class size which are a little smaller and less significant than when only the linear enrolment term is included.

Table 3 shows that French class size has a marginally significant negative coefficient in the model for marks in English, when enrolment is included as control. It is also close to being significant at the 5% level in the regressions for average marks and science marks when controlling for enrolment, and in regressions for English marks without control for enrolment. In all other regressions it is clearly insignificant. Controlling for enrolment, the point estimates for English marks are about 40% of the corresponding point estimates for French marks in Table 1, whereas the point estimates for science and average marks are respectively 27-28% and 18-19% of the estimated effects on French marks. At least some of the effect on English marks may be due to spill-over effects from the class-size effect on proficiency in French, as mentioned. However, the effect on English marks is quite large, and the spill-over interpretation is less obvious for science. We cannot therefore exclude the possibility that part of the estimated class-size effects in Table 1 is due to selection mechanisms causing a positive correlation between small class size and a large fraction of ambitious or high-ability students, especially in the learning of foreign languages.[10]

[Table 3 about here]

Two strategies are used to deal with this possible selection problem. First, other marks are included as additional controls in regressions of French marks on French class size. Second, we estimate differences-in-marks models, in which the differences between French and other marks are regressed on class size in French and controls.

[10] Aaronson et al. (2007) investigate the effects of another aspect of school quality, namely teacher quality, and find statistically and educationally significant effects of math teachers on math test scores. They also find significant effects of English teachers on math test scores (and of math teachers on English test scores). As here, they are not able to distinguish to what extent this is due to spill-over effects between subjects and to what extent it is due to sorting mechanisms.

Controlling for marks in other subjects

Table 4 shows estimated effects of class size on examination marks in French when examination marks in other subjects are included as additional controls. In this case enrolment (and enrolment squared) is clearly insignificant, and its inclusion is not important for the estimated class-size effects. This is not surprising, since controlling for marks in other subjects should take account of possible correlation between enrolment and student ability. Consequently, Table 4 only shows results without control for enrolment. The table shows results with and without controls for individual background characteristics, and with and without control for examination marks in science. The reason why we show results without controlling for science marks is that many (about 200) students do not take this examination (see also Table 3). Table 4 shows highly significant class-size effects. The point estimates are numerically a little smaller (-0.045 to -0.050) than when marks in other subjects are not included as controls (in which case the estimates are in the range of -0.054 to -0.061; see Table 1), but the standard errors are also smaller, so that t-ratios are still about 4. Not surprisingly, marks in other subjects are highly significant in the estimations for French marks; including these extra controls improves the fit considerably and reduces the significance of individual family background controls. For instance, in the first estimation of Table 4, the t-ratios of marks in other subjects are between 5 and 14, the largest being for Danish spelling and English and the smallest for math, and R^2-within is 0.43 (compared to 0.10 in Table 1).

[Table 4 about here]

The fact that controlling for marks in science reduces the point estimates by about 10% is less an effect of controlling for science *per se* than a selection effect caused by excluding the relatively weak group of students who do not take the examination in science. Thus, estimating models for French marks while controlling for marks in all other subjects except science (as in the last two rows of Table 4), but restricting the sample to the 6195 observations which do have science examination marks, gives estimates of class-size effects of -0.0462 and -0.0460 (which are very close to the estimates in the first two rows of Table 4).

Estimating differences-in-marks models

Table 5 shows the results of estimating differences-in-marks models in which the difference between examination marks in French and another subject is regressed on French class size and controls. Again, enrolment is clearly insignificant, so that only results not controlling for this variable are shown. Given the definition of the LHS variables in the estimations in Table 5, it is not surprising that the parameter estimates in Table 5 are approximately equal to the difference between the corresponding estimates in Tables 1 and 3; the only reason why this relationship is not exact is that a few students who have examination marks in French do not have marks in other subjects.[11] Most differences-in-marks estimates in Table 5 are about -0.05 (with t-ratios between 3.5 and 4.2) as in Table 4, but when marks in English or science are subtracted from French marks the estimates become numerically a little smaller, namely about -0.04 (with t-ratios of about 3).

[Table 5 about here]

The fact that controlling for (or differencing with respect to) examination marks in other subjects reduces the estimated effect of French class size on French marks may be interpreted as spill-over effects from French skills to marks in other subjects, or as a sign of selection of high-

[11] Thus, the sample size in the estimations of Table 1 is larger than the sample sizes in Tables 3 and 5 (except for the average of marks in all other subjects which are calculated from the marks observed for each individual).

ability students into small French classes (see the discussion above). In the first case one should not control for marks in other subjects, and rely instead on the estimates in Tables 1 or 2. In the second case, the estimates in Tables 4 and 5 may be closer to the truth. It is of course possible that both spill-over and selection effects are important. Nevertheless, even with the selection interpretation, the estimated class-size effects correcting for this selection remain highly significant.

Who benefits from a small class size?

This subsection considers the extent to which class-size effects depend on student characteristics such as gender, family background and academic ability. The most significant explanatory variables in the regressions discussed above are the gender dummy, parental education variables, and especially the variables for marks in other subjects. Table 6 shows results of estimation from ten different regressions of French examination marks on French class size and the full set of individual background controls, and also an interaction term between class size and some other variable. The variables in interaction with class size are: the female dummy, a dummy for parents having a high level of education (at least one parent has a long further or a higher education), a dummy for the mother having an academic education (either upper secondary school or a further or higher education), a dummy for average marks in other subjects being at least 9 (i.e., above average), and average marks in other subjects (treated as a continuous variable).[12] For each interaction variable, the model is estimated in two versions, with and without six controls for examination marks in the other subjects (i.e., English, Danish (three variables) and math (two variables)).[13]

[Table 6 about here]

[12] When the variable interacting with class size is not a linear combination of other controls, it is included as an extra control. Enrolment is included as control if its t-ratio is above 0.8, but this is not important for the results shown.

[13] Examination marks in science are not controlled for since we would then lose about 200 observations, but the results are very similar with science marks included.

The first two regressions in the table show that class-size effects are significantly (34-38%) smaller for females than for males, and the effects for males are 24-29% larger than in the corresponding estimations without interaction terms (the second estimation in Table 1 and the third estimation in Table 4, respectively). When marks in other subjects are not controlled for, the estimated class-size effects are significantly smaller for children whose parents have a high level of education (see the third and fifth estimation in Table 6). When marks in other subjects are controlled for, this difference becomes insignificant (see the fourth and sixth estimation). This is not surprising, since examination marks in other subjects are clearly intervening variables regarding parental background. Thus, when academic ability is controlled for, class-size effects do not depend on parental education, while they remain very different for boys and girls.

The last four regressions in Table 6 show that class-size effects are significantly smaller for students with higher examination marks in other subjects. Upon including detailed controls for marks in other subjects, the t-ratios of interaction terms are about 3. When average marks in other subjects are above average (i.e., at least 9), class-size effects are 34-44% smaller than when average marks are below 9. To interpret the results of the last two regressions, in which the interaction term is defined in terms of the continuous variable for average marks in other subjects, we calculate the class-size effect for different values of this variable. The result is shown in the first column of Table 8. For students in the estimation sample, average marks in other subjects vary between 4.7 and 11.7, with an average of 8.8.[14] The first column of Table 8 shows class-size effects for the last estimation in Table 6 calculated for different percentiles in the distribution of average marks in other subjects. The differences in estimated effects are large. For students with low academic ability (1st to 5th percentile in the distribution of average marks) the

[14] The average for all students is about 8.0, so that students taking the examination in French have on average rather high marks in other subjects.

class-size effect on French examination marks is 0.07-0.08 (numerically), whereas it is only 0.02-0.03 for high-ability students (95th to 99th percentile). The class-size effect is about twice as large at the 10th percentile as at the 90th percentile.

[Table 7 about here]

Table 7 shows estimation results when two class-size interaction effects are included simultaneously in the model: interaction between class size and the female dummy, and interaction between class size and one more variable.[15] The second interaction is with respect to one of the last four variables used in interaction terms in Table 6: the two dummies for parental education, and the two variables for average examination marks in other subjects. Again, class size is highly significant. Interaction with the female dummy is significant at the 10% level in all regressions, and at the 5% level in regressions not including marks in other subjects; point estimates are approximately equal to the estimates in the first two rows of Table 6. Interaction effects between class size and parental education are similar to the corresponding estimates of Table 6; the same is true for interaction effects with respect to average marks in other subjects. Thus, interactions with respect to both gender and average marks in other subjects are significant when included simultaneously. This shows that the significance of the gender interaction effect is not simply due to females having higher marks than males. The class-size effects implied by the estimates of the last row of Table 7 with the continuous variable for average marks in other subjects are shown in the last two columns in Table 8.

[Table 8 about here]

[15] Interaction between the female dummy and the other variable interacted with class size is also included.

6. Comparisons

When controlling for individual background variables, the estimates of average class-size effects in the present study are in the range from -0.0392 (Table 5, the difference between French and English marks) to -0.0645 (Table 2). To compare with other published estimates of the effect of class size, it is useful to divide the estimates by the standard deviation of the examination marks. The standard deviation is 1.97 in the distribution of marks of individual students; it is 1.05 in the distribution of average marks of school/year cells. Comparisons are made below with the results from the STAR experiment in the US (see, e.g., Finn and Achilles, 1990, Krueger, 1999, and Nye et al., 2000), in which students and teachers were randomly assigned to small classes (13-17 students) or regular-size classes (22-25 students) with an average difference of 8 students in class size. For a class-size reduction of 8 students, the results of the present paper suggest an increase in French examination marks in the range 0.31-0.52, corresponding to 0.16-0.26 standard deviations in the individual distribution of test scores, and 0.30-0.49 standard deviations in the distribution of class means.

The conclusion of Krueger (1999) from the STAR experiment of reducing (basic) class size in grades up to grade 3 is that a reduction of 8 students per class increases average test scores (reading, word recognition, and math, Stanford Achievement Test) by 0.19-0.28 standard deviations in the distribution of students' individual test scores; see Krueger (1999, Table V, and page 514). The estimated effect is largest for 1st grade and smallest for 3rd grade. Thus, the effect sizes of 0.16-0.26 in the present paper are very close to the estimates based on the STAR experiment.

Using Maimonides' maximum class-size rule on an Israeli dataset with an average class size of 32 and a maximum of 40, Angrist and Lavy (1999) find significant class-size effects on reading and math test scores for 5th graders. The estimated class-size coefficients are -0.260 to

-0.687 for reading scores, and -0.261 to -0.596 for math scores (smallest for full sample estimates and largest for discontinuity sample estimates with dummy instrumental variables; see Tables IV and VI in their paper). Upon reducing the number of students by 8, this corresponds to effect sizes of 0.27-0.67 standard deviations for reading scores and 0.22-0.47 standard deviations for math scores, where standard deviations are for the distributions of class average test scores. These effect size ranges are close to the corresponding estimates (0.30-0.49) from the present work. However, Angrist and Lavy do not find significant effects for 4th and 3rd graders (and the point estimates are much smaller). Hoxby (2000) applies maximum class-size rules and random population variation to estimate class-size effects on test scores using Connecticut elementary school data, but does not find any significant effects.

Using a large panel dataset for students in Texas, Rivkin et al. (2005) estimate effects of class size on test score gains in math and reading, controlling for student and school (or school-by-year) fixed effects. They find highly significant negative effects for 4th and 5th grade, and significant effects for 6th grade in math; effects for 7th grade (and for 6th grade in reading) are insignificant. Estimates for 4th grade indicate that a class-size reduction of 8 increases test scores in math and reading by about 0.12 and 0.10 standard deviations respectively in the distribution of students' individual test score gains. The corresponding effect-size estimates for 5th grade are 0.10 for math and 0.04 for reading. These effect-size estimates are considerably smaller than those found in the present paper, but this is to be expected since the dependent variable is test score *gains* (from one grade to the next).

The results of the present paper when including interaction terms between on the one hand class size and on the other parental education, gender, and academic ability indicate that boys benefit more from small classes than girls, and that low-ability students benefit more than high-ability students. The class-size effect is also larger for students whose parents have a low level of education, but the difference is not significant when academic ability (measured by marks in other subjects) is controlled for. The results of Summers and Wolfe (1977), Krueger (1999), Angrist and Lavy (1999), Heinesen and Graversen (2005) and Browning and Heinesen (2007) also indicate that increasing school resources may have larger effects for students from disadvantaged backgrounds, and results in Browning and Heinesen (2007) and Dustmann et al. (2003) indicate larger effects for boys than for girls. However, the differences between subgroups seem more significant in the present paper than in earlier studies, and estimates of class-size effects as functions of academic ability (measured by marks in other subjects) are not found in earlier studies.

7. Conclusion

The main contribution of this paper is the identification of effects of class size, using variation in the size of subject-specific classes of alternative subjects between which students must choose. This variation is not likely to be influenced by the endogeneity problems which affect the existing literature focussing on variation in the size of basic classes (i.e., classes in which most topics are taught), especially the selection response of parents to poor school quality, for instance large class size.

In Danish schools students have to choose between German and French at 7th grade, so that these subjects are taught in classes which are different from the basic classes in which other subjects are taught. By exploiting within-school variation over time in the size of French classes, which reflects random variation in the number of students choosing to learn French instead of German, the present paper finds consistently positive effects on French examination marks of reducing French class size. The analysis covers class-size variations between 1 and 20 students. Within this range the linearity of class-size effects is not rejected, although there is some indication of negative effects of reducing class size below 5 students.

French class size has a small, but statistically significant, negative effect on English examination marks, whereas there are no significant effects on marks in Danish or math; effects on science marks are negative and close to being significant. The effect on English marks may be explained by spill-over effects (if a small French class size improves French language skills this may spill over to English marks via a better foundation for learning other foreign languages) or by selection effects (such that small classes are associated with students having higher average ability or being more ambitious in terms of learning foreign languages).

If there is no selection, the class-size estimates of Tables 1 and 2 are the relevant estimates. These estimates are about -0.06 in value, with t-ratios of about 4. If, on the other hand, the correlation between French class size and marks in other subjects (especially English) is caused by selection effects, the estimates of Tables 4 and 5 are the relevant ones. These estimates are in the range -0.04 to -0.05, with t-ratios of 3-4. Dividing the estimates of class-size effects in the present paper by the standard deviation of the examination marks, the estimates are of about the same magnitude as the estimates in Krueger (1999) and Angrist and Lavy (1999).

Identifying class-size effects by variation in subject-specific classes makes it possible to estimate class-size effects as functions of academic ability measured by marks in other subjects. The results of the present paper when including interaction terms between on the one hand class size and on the other parental education, gender, and academic ability indicate that reducing class size may promote intergenerational mobility and have positive effects in promoting equality. They also indicate that boys benefit more from small classes than girls, and that low-ability students benefit more than high-ability students.

An important issue is the extent to which the results of the present paper apply in other contexts. First, class-size effects on marks or test scores in other subjects than French may be different. Small class size might be important when students learn a foreign language with a structure and pronunciation which is very different from their mother tongue, whereas it might be less important for learning math, for example. Second, class-size effects are estimated at 9th grade, and effects at earlier grades might be different. Even though the effects estimated in this paper are of the same order of magnitude as those found in the STAR experiment and by Angrist and Lavy (1999), this could be due to larger effects of class size in lower grades, for instance, *and* larger effects for learning foreign languages than other subjects. Third, variation was considered in subject-specific classes, but class-size effects might be different in basic classes (in which most subjects are taught) because basic classes are more important for establishing social relations among students, which may also be important for learning. In particular, there might be negative effects on social relations from reducing class size below a certain point – effects that could work against possible positive effects of reducing class size. Fourth, we observe large and approximately linear effects of class size below 20 pupils. We may also expect negative effects of increasing class size above 20. According to Lazear's (2001) theoretical model of educational production, non-linear 'disruption effects' may cause the marginal effect of increasing class size to decrease, but such non-linearity is not likely to be important below a class size of about 30 students.

This discussion of 'external validity' highlights the fact that much more research is needed on the effects of class size and the complex mechanisms through which they operate.

[Appendix 1. Table A1 about here]

Appendix 2. Further tests of linearity

This appendix reports results of F tests of linearity of class-size effects when the number of dummies for class size is smaller than 19. For instance, non-linearity was tested by replacing the class-size variable, *n*, in equation (1) by nine dummy variables for class-size intervals of two. Denoting dummy variables for class size by d_j as in Section 5, the redefined dummies are:

$$d_j = 1 \text{ if } 2j-1 \le n \le 2j, \quad d_j = 0 \text{ otherwise}, \quad j = 1,\ldots,9 \tag{3}$$

The reference category is defined as class size equal to 19 or 20. Denoting again the coefficients of these redefined dummy variables d_j by α_j, linearity in this model is now implied by the following eight linear restrictions:

$$\frac{\alpha_j - \alpha_{j+1}}{\overline{n}_j - \overline{n}_{j+1}} - \frac{\alpha_9}{\overline{n}_9 - \overline{n}_{10}} = 0, \quad j = 1,\ldots,8 \tag{4}$$

where $\overline{n}_j$ denotes average class size when $d_j = 1$, and $\overline{n}_{10}$ denotes average class size for the reference category ($n = 19$ or $n = 20$). The denominators on the LHS in (4) take account of the fact that the average class size when $d_{j+1} = 1$ is not exactly two pupils more than when $d_j = 1$ as a result of different numbers of observations for each value of class size (although this is not important for the outcome of the tests in this case). Upon estimating this model with the nine dummy variables and testing the eight restrictions in (4), linearity cannot be rejected; the P values of the F(8, 374) test statistic are in the range 0.87-0.90 including individual controls, and 0.44-0.51 excluding these controls. Reducing the number of dummies for class size further, linearity is still not rejected; including four dummies (for class-size ranges 1-4, 5-8, 9-12, and 13-16, with 17-20 as reference category), and testing the three restrictions similar to (4), the P values are 0.24-0.50.

References

Aaronson, D., Barrow, L. and Sander, W. (2007), Teachers and student achievement in Chicago public high schools, *Journal of Labor Economics 25* (1), 95-135.

Angrist, J.D. and Lavy, V. (1999), Using Maimonides' rule to estimate the effect of class size on scholastic achievement, *Quarterly Journal of Economics 114* (2), 533-575.

Browning, M. and Heinesen, E. (2007), Class size, teacher hours and educational attainment, *Scandinavian Journal of Economics*, forthcoming.

Card, D. and Krueger, A.B. (1996), *Labor market effects of school quality: Theory and evidence*, in G. Burtless (ed.), *Does money matter? The effect of school resources on student achievement and adult success,* Brookings Institution Press, Washington, D.C.

Dearden, L., Ferri, J. and Meghir, C. (2002), The effect of school quality on educational attainment and wages, *Review of Economics and Statistics 84* (1), 1-20.

Dustmann, C., Rajah, N. and van Soest, A. (2003), Class size, education, and wages, *Economic Journal 113*, F99-F120.

Finn, J.D. and Achilles, C.M. (1990), Answers and questions about class size: a statewide experiment, *American Educational Research Journal 27* (3), 557-577.

Hanushek, E.A. (1996), School resources and student performance, in G. Burtless (ed.), *Does money matter? The effect of school resources on student achievement and adult success,* Brookings Institution Press, Washington, D.C.

Hanushek, E.A. (1999), Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects, *Educational Evaluation and Policy Analysis 21* (2), 143-163.

Hedges, L.V. and Greenwald, R. (1996), Have times changed? The relation between school resources and student performance, in G. Burtless (ed.), *Does money matter? The effect of school resources on student achievement and adult success*, Brookings Institution Press, Washington, D.C.

Heinesen, E. and Graversen, B.K. (2005), The effect of school resources on educational attainment: Evidence from Denmark, *Bulletin of Economic Research 57* (2), 109-143.

Hoxby, C. (2000), The effects of class size on student achievement: New evidence from population variation, *Quarterly Journal of Economics 115*, 1239-1285.

Krueger, A.B. (1999), Experimental estimates of educational production functions, *Quarterly Journal of Economics 114* (2), 497-532.

Krueger, A.B. (2003), Economic considerations and class size, *Economic Journal 113*, F34-F63.

Lazear, E.P. (2001), Educational production, *Quarterly Journal of Economics 116* (3), 777-803.

Nye, B., Hedges, L.V., and Konstantopoulos, S. (2000), The effect of small classes on academic achievement: The results of the Tennessee class size experiment, *American Educational Research Journal 37*, 123-151.

Rivkin, S.G., Hanushek, E.A. and Kain, J.F. (2005), Teachers, schools, and academic achievement, *Econometrica 73* (2), 417-458.

Summers, A.A., and Wolfe, B.L. (1977), Do schools make a difference?, *American Economic Review 67* (4), 639-652.

West, M.R., and Wössmann, L. (2006), Which school systems sort weaker students into smaller classes? International evidence, *European Journal of Political Economy 22* (4), 944-968.

Table 1. Effect of class size in French on examination marks in French; results from within-school regressions

Class-size estimate	SE of estimate	Enrolment	Enrolment squared	Individual controls
-.0593	.0151	Yes	Yes	Yes
-.0607	.0148	Yes	No	Yes
-.0559	.0143	No	No	Yes
-.0582	.0144	Yes	Yes	No
-.0595	.0143	Yes	No	No
-.0543	.0137	No	No	No

Note: The number of observations is 6507. Robust standard errors are corrected for clustering within schools.

Table 2. Effect of class size in French on examination marks in French; results from within-school regressions including a dummy variable for class size smaller than 5

Class-size estimate	SE	Dummy variable estimate	SE	Enrolment	Enrolment squared	Individual controls
-.0631	.0163	-.2300	.1732	Yes	Yes	Yes
-.0645	.0162	-.2166	.1721	Yes	No	Yes
-.0591	.0156	-.1997	.1714	No	No	Yes
-.0648	.0156	-.4066	.1840	Yes	Yes	No
-.0662	.0156	-.3930	.1825	Yes	No	No
-.0602	.0150	-.3741	.1833	No	No	No

Note. Robust standard errors corrected for clustering within schools.

Table 3. Effect of class size in French on examination marks in other subjects, within-school regressions

Subject	Class-size estimate	SE of estimate	Enrolment as control	Individual controls	Number of observations
Average, all subjects	-.0116	.0060	Yes	Yes	6507
	-.0064	.0057	No	Yes	6507
	-.0109	.0066	Yes	No	6507
	-.0047	.0062	No	No	6507
English	-.0233	.0106	Yes	Yes	6484
	-.0170	.0103	No	Yes	6484
	-.0240	.0108	Yes	No	6484
	-.0169	.0106	No	No	6484
Danish (oral)	-.0150	.0100	Yes	Yes	6496
	-.0046	.0096	No	Yes	6496
	-.0127	.0102	Yes	No	6496
	-.0015	.0097	No	No	6496
Danish (written)	-.0041	.0077	Yes	Yes	6463
	-.0036	.0071	No	Yes	6463
	-.0033	.0084	Yes	No	6463
	-.0014	.0078	No	No	6463
Danish (spelling)	-.0084	.0088	Yes	Yes	6475
	-.0042	.0082	No	Yes	6475
	-.0080	.0090	Yes	No	6475
	-.0026	.0084	No	No	6475
Math (oral)	-.0067	.0088	Yes	Yes	6481
	-.0005	.0085	No	Yes	6481
	-.0077	.0088	Yes	No	6481
	-.0003	.0086	No	No	6481
Math (written)	-.0091	.0089	Yes	Yes	6467
	-.0059	.0080	No	Yes	6467
	-.0082	.0094	Yes	No	6467
	-.0033	.0085	No	No	6467
Science	-.0172	.0094	Yes	Yes	6288
	-.0108	.0093	No	Yes	6288
	-.0158	.0099	Yes	No	6288
	-.0084	.0097	No	No	6288

Note: Robust standard errors corrected for clustering within schools.

Table 4. Effect of class size in French on examination marks in French, controlling for marks in other subjects, within-school regressions

Class-size estimate	SE of estimate	Control for science marks	Individual controls	Number of observations
-.0453	.0126	Yes	Yes	6195
-.0454	.0119	Yes	No	6195
-.0501	.0124	No	Yes	6393
-.0505	.0117	No	No	6393

Note: Robust standard errors corrected for clustering within schools.

Table 5. Effect of class size in French on the difference between examination marks in French and other subjects, within-school regressions

Subject whose marks are subtracted from French marks	Class-size estimate	SE of estimate	Individual controls	Number of observations
Average, all subjects	-.0495	.0123	Yes	6507
	-.0496	.0118	No	6507
English	-.0392	.0131	Yes	6484
	-.0378	.0129	No	6484
Danish (oral)	-.0520	.0140	Yes	6496
	-.0533	.0132	No	6496
Danish (written)	-.0524	.0140	Yes	6463
	-.0526	.0132	No	6463
Danish (spelling)	-.0519	.0138	Yes	6475
	-.0521	.0128	No	6475
Math (oral)	-.0549	.0145	Yes	6481
	-.0532	.0140	No	6481
Math (written)	-.0496	.0142	Yes	6467
	-.0507	.0141	No	6467
Science	-.0414	.0150	Yes	6288
	-.0418	.0147	No	6288

Note: Robust standard errors corrected for clustering within schools.

Table 6. Effect of French class size on French examination marks; within-school regressions including individual background controls, and interactions between class size and gender, parental education or average marks in other subjects

Variable interacting with class size	Class-size estimate	SE	Class-size-interaction estimate	SE	Controlling for marks in other subjects	Number of observa-tions
Female	-.0785	.0169	.0303	.0127	No	6507
	-.0623	.0140	.0212	.0104	Yes	6393
Parents high education	-.0737	.0161	.0247	.0120	No	6507
	-.0532	.0134	.0059	.0099	Yes	6393
Mother academic education	-.0739	.0165	.0251	.0113	No	6507
	-.0562	.0137	.0114	.0095	Yes	6393
Average of other marks ≥ 9	-.0605	.0138	.0204	.0102	No	6507
	-.0622	.0129	.0273	.0090	Yes	6393
Average of other marks	-.1630	.0424	.0131	.0048	No	6507
	-.1707	.0429	.0139	.0048	Yes	6393

Note: Robust standard errors corrected for clustering within schools.

Table 7. Effect of French class size on French examination marks; within-school regressions including individual background controls, and interactions between class size and gender, and between class size and parental education or average marks in other subjects

Variable interacted with class size (and gender)	Class-size estimate	SE	Class-size-female interaction	SE	Class-size-other interaction	SE	Controlling for marks in other subjects
Parents high education	-.0911	.0181	.0287	.0127	.0262	.0120	No
	-.0653	.0151	.0203	.0104	.0068	.0099	Yes
Mother academic educat.	-.0919	.0186	.0284	.0128	.0277	.0113	No
	-.0687	.0155	.0203	.0105	.0129	.0095	Yes
Average other marks ≥ 9	-.0721	.0154	.0207	.0115	.0199	.0102	No
	-.0726	.0142	.0187	.0104	.0268	.0090	Yes
Average of other marks	-.1719	.0427	.0191	.0107	.0128	.0048	No
	-.1791	.0431	.0179	.0104	.0137	.0048	Yes

Note: Robust standard errors corrected for clustering within schools.

Table 8. Effect of French class size on French examination marks by percentiles in the distribution of average examination marks in other subjects (based on last estimations in Tables 6 and 7)

Average marks in other subjects		Class-size effects in French		
Percentile	Average marks	All (Table 6)	Males (Table 7)	Females (Table 7)
1	6.2	-0.0845	-0.0942	-0.0763
5	7.0	-0.0734	-0.0832	-0.0653
10	7.4	-0.0678	-0.0777	-0.0598
20	7.9	-0.0609	-0.0709	-0.0530
30	8.3	-0.0553	-0.0654	-0.0475
40	8.6	-0.0512	-0.0613	-0.0434
50	8.9	-0.0470	-0.0572	-0.0393
60	9.1	-0.0442	-0.0544	-0.0365
70	9.2	-0.0428	-0.0531	-0.0352
80	9.7	-0.0359	-0.0462	-0.0283
90	10.0	-0.0317	-0.0421	-0.0242
95	10.4	-0.0261	-0.0366	-0.0187
99	11.0	-0.0178	-0.0284	-0.0105

Appendix 1. Table A1. Descriptive statistics and within-school estimates for the full set of controls. The estimation result corresponds to the first row of Table 1.

	Descriptive statistics				Estimation result	
Variable	Mean	Std. Dev.	Min	Max	Coeff.	t-ratio
French examination marks	7.854	1.970	0	13		
French class size	12.435	4.413	1	20	-0.059	-3.93
Year 2002	0.322	0.467	0	1	-0.045	-0.60
Year 2003	0.327	0.469	0	1	-0.045	-0.54
Female	0.598	0.490	0	1	0.762	14.03
Immigrant	0.039	0.195	0	1	0.151	0.82
Second generation immigrant	0.048	0.214	0	1	0.269	1.92
Number of siblings	1.074	0.866	0	7	-0.077	-1.99
Has younger siblings	0.457	0.498	0	1	0.125	2.10
Broken family	0.305	0.460	0	1	-0.243	-4.36
Mother not in the register	0.009	0.094	0	1	0.418	1.36
Father not in the register	0.052	0.223	0	1	0.443	3.26
Mother teenager (at time of birth)	0.028	0.165	0	1	-0.386	-2.44
Mother upper secondary school	0.060	0.238	0	1	0.583	5.09
Father upper secondary school	0.061	0.240	0	1	0.340	3.30
Mother vocational education	0.258	0.437	0	1	0.090	1.24
Father vocational education	0.304	0.460	0	1	0.111	1.53
Mother short further education	0.063	0.242	0	1	0.647	6.02
Father short further education	0.049	0.217	0	1	0.260	2.09
Mother long further education	0.313	0.464	0	1	0.584	7.59
Father long further education	0.166	0.372	0	1	0.425	5.15
Mother higher education	0.108	0.310	0	1	0.989	9.94
Father higher education	0.198	0.398	0	1	0.607	6.51
Mother’s education unknown	0.021	0.142	0	1	0.297	1.49
Father’s education unknown	0.020	0.139	0	1	0.003	0.01
Mother self-employed	0.045	0.207	0	1	0.015	0.11
Father self-employed	0.102	0.303	0	1	0.010	0.11
Mother student	0.016	0.125	0	1	-0.178	-0.93
Father student	0.003	0.058	0	1	-0.051	-0.11
Mother receives social assistance	0.035	0.183	0	1	-0.221	-1.20
Father receives social assistance	0.016	0.125	0	1	-0.089	-0.36
Mother not in the labour market	0.044	0.205	0	1	0.048	0.36
Father not in the labour market	0.043	0.203	0	1	0.106	0.86
Log family income from employment	3.448	1.039	-9.407	5.705	0.000	-0.01
Family income from employment zero	0.071	0.257	0	1	-0.010	-0.06
Mother’s degree of unemployment	4.090	14.478	0	100	-0.002	-1.02
Father’s degree of unemployment	3.078	13.181	0	100	-0.001	-0.57
Number of rooms per person	1.187	0.438	0	4.750	0.067	0.97
Number of rooms per person unknown	0.004	0.062	0	1	0.016	0.04
Share of students without French or German	0.113	0.099	0	0.708	-0.640	-1.12
Number of students (at 9th grade)	41.840	14.031	3	106	-0.005	-0.22
Number of students squared/100	19.474	13.840	0.09	112.360	0.012	0.53
Constant					7.555	14.36

Note: t-ratios are based on robust standard errors corrected for clustering within school. The number of observations is 6507.

Figure 1. Histogram of examination marks

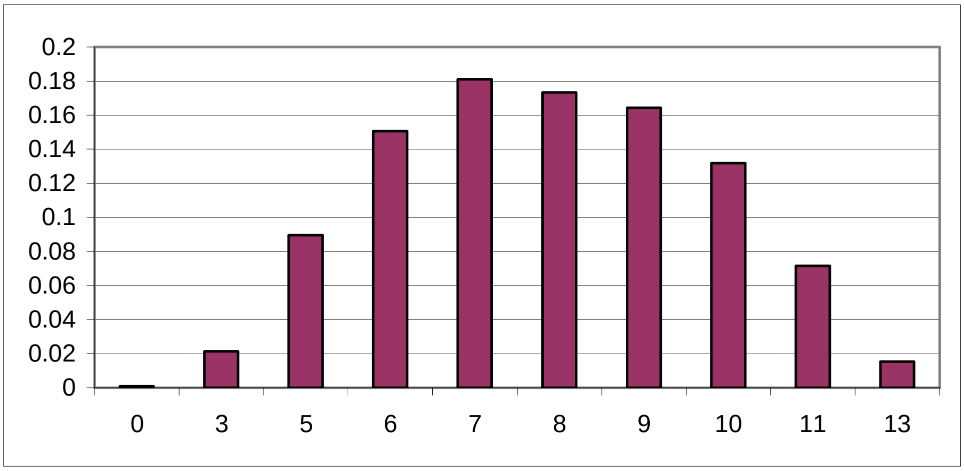

Figure 2. Number of students by class size in French classes

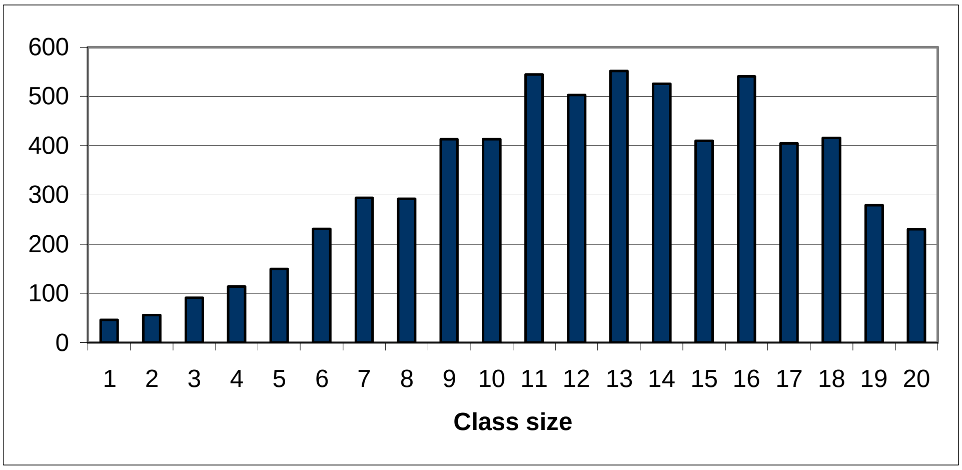

Figure 3. Average marks and residuals from school fixed-effects regressions of marks on (a) time dummies, and (b) time dummies and other controls, by class size, in class-size intervals of two