Essays in the economics of education
Levin, J.D.

Citation for published version (APA):
Levin, J. D. (2002). Essays in the economics of education

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 1

Class size, peers and educational production

1.1 Introduction

As in many other countries, the causal effect of class size on pupils’ school performance is an important issue in the ongoing educational debate in the Netherlands. In 1997, the Dutch government committed to undertake a major investment to reduce average class size in primary schools. Although opinion leaders and politicians were easily persuaded that this was a good investment, the measure was not backed by convincing empirical evidence pertaining to the Dutch situation. In the process of policy preparation one study was published supporting the reduction of class size. As we will show in this chapter, the results of this study should be questioned.

Traditionally, the study of the effect of class size on school performance has been dominated by educational researchers. Only recently have economists entered this field. A notable early exception is the influential survey article by Hanushek (1986) who, after reviewing a large number of studies, concludes that there is no positive effect of schooling expenditures on pupils’ school performance.

The key methodological problem to assess the causal effect of class size on pupils’ school performance is that assignment of pupils to classes of different sizes is unlikely to be random. Consequently, differences in performance between pupils from small and large classes can be caused by factors other than differences in class size. Especially, two mechanisms are likely to play a role. Firstly, parents may choose the school to which they send their children based partly on class size. It is likely that parents who care more about their children’s achievement in school are more inclined to make sure their children are in schools with smaller average class sizes and/or in smaller classes within a school. If that is the case and the degree to which parents care about their children’s achievement has an independent effect on children’s school performance, but is unobserved, then the estimated coefficient on class size will be biased downwards. Secondly, if there is more than one class at a particular grade level, the size and composition of each class depend on choices made by the administration of a school. The school might decide to form larger and smaller classes at the same grade level and assign weaker pupils to the smaller ones. If controlling for this assignment process is not possible, the coefficient on class size will be biased upwards and may even be positive (i.e. it will appear that smaller classes are associated with lower achievement). Hence, an uncorrected estimate of the class size effect can be biased upwards or downwards depending on which of the two sources of bias plays the larger role in a particular setting.

The most attractive way to encompass these problems is by running a real field experiment. This was done in the state of Tennessee, where the Student/Teacher Achievement Ratio (STAR) experiment was held. Students were randomly assigned to classes of different sizes. But even in that case, serious problems such as reassignment and
attrition, which possibly biased the results, occurred. In a careful reanalysis of the STAR experiment, Krueger (1999) concludes that in this experiment class size reduction significantly boosted student performance. In a follow-up analysis, Krueger and Whitmore (2001) analyze the long-term effects of the intervention and find that these are present. Of course, organizing a field experiment is generally not within the reach of researchers. Therefore, to collect information applicable to other settings than those in Tennessee, different approaches have to be applied.

In a recent paper Angrist and Lavy (1999) propose a method which is closely related to the real experiment type of study. The authors construct an instrument for class size using a rule contained in medieval Jewish law. Maimonides’ Rule (named after the rabbinical scholar who scribed it) dictates the number of teachers to be allocated to a given class based on the number of participating students. The authors show the rule to produce a functional relationship between average class size (as predicted by the rule) and enrollment that is non-monotonic and discontinuous, thereby giving support to the validity of the former as an instrument for actual average class size.¹

In the current paper we apply a methodology close to the one applied by Angrist and Lavy, but in the context of the Dutch educational system. The instrument used here is based on the rules that link total enrollment in the school to the number of teachers a school receives funding for. These rules also cause discontinuities in the relation between enrollment and average class size and pupil-teacher ratio. These discontinuities are clearly smaller than those in Israel, but still lead to a considerable degree of exogenous variation of class size. The dataset used here contains individual information on more than 12,000 pupils over three different grade levels (4, 6, and 8).

Corrected for a large number of observables and for endogeneity bias, our estimation results indicate that pupils in smaller classes do not perform better (and sometimes even worse) than pupils in larger classes. In a subsequent analysis we attempt to explain this finding, by decomposing the class size effect into two underlying effects. Reduction of class size causes two changes. One change is that the number of pupils per teacher decreases. This allows the teacher to spend more teaching time per pupil, which seems likely to improve pupils’ achievement. A second change caused by reducing class size is that the number of classmates one has is reduced. If pupils learn not only from their teacher but also from their classmates, this change may harm pupils’ achievement. Social cognitive learning theory strongly suggests that pupils’ achievement may benefit from a larger number of classmates with similar levels of cognitive ability. To disentangle the two effects, we construct a new variable, which measures for each pupil in the sample the number of classmates with similar IQ. Inclusion of this variable in the analysis reveals that the number of similar pupils has, indeed, a significantly positive impact on performance. Moreover the class size effect has now, in most cases, the proper negative sign. It is important to stress that our measure of similar pupils counts the absolute number of similar pupils within a class. This variable has an expected value that is increasing in class size, which differs from usual measures of class homogeneity.

¹ More precisely, the relationship allows one to isolate the exogenous variation in class size. Isolating the exogenous portion of class size and estimating its effect on achievement can be considered analogous to running a field experiment (also known as a quasi-experiment). Support for the validity of such an instrument can be found in the literature on regression-discontinuity design (see Campbell and Stanley, 1963).
The organization of the remainder of this chapter is the following. In the next section we summarize in some more detail the methods and results of recent related studies. We focus on two studies in particular. The first is the aforementioned article by Angrist and Lavy (1999) and the second by Bosker and Hox (1996), the latter being a Dutch study that played an important role in persuading Dutch politicians that reducing class size is an effective means to boost scholastic achievement. Their study uses the same dataset as we will use for our own analyses. In Section 1.3 we describe the dataset and the construction of variables. The section also contains an account of the creation of our instrumental variable. Section 1.4 describes, discusses and explains the empirical findings. Section 1.5 is devoted to the robustness of the findings and discusses the outcomes of different specifications and different sub-samples. Section 1.6 summarizes and concludes.

1.2 Related studies

Class sizes in public schools in Israel are partly determined by a rule proposed by the 12th century scholar Maimonides. According to this rule up to 40 pupils may be put in one class. Should there be more than 40 pupils, an additional class will be created. Consequently, if there are 40 pupils at a given grade level, the average class size equals 40. With one additional pupil the average class size drops to 20.5. When enrollment rises further average class size increases again to 40 when enrollment equals 80. The 81st pupil causes the creation of a third class, so that the average class size drops again to 27.

Applying Maimonides' rule leads to a discontinuous relation between enrollment and class size. In practice, this rule is not strictly followed; actual class sizes differ from those predicted by Maimonides' rule. But as long as actual class size is partly determined by this rule, the rule serves as a credible source of exogenous variation of actual class size. As such it can be used to create an instrumental variable which purges actual class size of its endogenous component. Class size according to the rule is a valid instrument if there is no effect of the instrument on pupils' school performance, other than the one caused by the relation between the instrument and actual class size.

Angrist and Lavy (1999) regress class average achievement scores on school characteristics and class size. This equation is estimated separately for 3rd, 4th and 5th grade classes in Israeli primary schools using both ordinary least squares (OLS) and instrumental variables (IV) techniques. The OLS results show a significantly positive effect of class size on reading and math scores when no control for socioeconomic status (SES) is included. The class size effects get much smaller, although some remain significantly positive when SES is controlled for. For the 4th and 5th grades, results change dramatically with the IV approach. In almost every specification the effect of class size on reading and math scores becomes significantly negative. This is not the case for 3rd grade pupils. As a possible explanation for this latter effect, Angrist and Lavy mention that the effect of small classes may operate cumulatively. It is likely that pupils who are in small classes in 5th grade were also in small classes in the 3rd and 4th grades. This explanation is supported by the finding that class size effects are larger for 5th than for 4th graders.

Some other recent studies also apply the IV approach to the effect of class size on pupils' school performance. Akerhielm (1995) uses the average class size for a given subject in the student's school and the 8th grade student enrollment in the school as instruments for
class size. She reports significantly negative effects of class size on school performance. Boozer and Rouse (1995) construct instruments from the maximum special education class size as dictated by state law. They too find significantly negative effects of class size on school performance. Hoxby (2000a) uses two independent identification strategies. The first exploits the random component of population variation, which causes exogenous variation in class size. Her second method is very close to the one developed by Angrist and Lavy and is based on the discontinuous changes in class size occurring when a small change in enrollment triggers a maximum or minimum class size rule. For both identification strategies Hoxby finds no effect of class size on achievement. The estimates are sufficiently precise to even rule out modest effects.

While the results of Akerhielm, and Boozer and Rouse are consistent with Krueger’s (1999) findings from the field experiment organized in Tennessee, Hoxby’s results are at odds with these findings. A possible explanation, suggested by Hoxby, is that in a field experiment participants are aware of being evaluated which may provide them with an incentive to make good use of the opportunities. Such explicit incentives are absent in the natural experiment approach employed in Hoxby’s study. This leaves unexplained, however, why the results obtained by Hoxby deviate from those reported in the other two IV studies. Although in this respect it should be pointed out that the exogenous variation exploited by Hoxby is more credible than the exogenous variation used in the other two studies.

In February 1996 the Dutch vice-minister of education asked an independent committee to advise her on the implication of smaller class sizes in primary education. The report includes an empirical study by Bosker and Hox (1996), using the so-called PRIMA data. Our own empirical analysis is based on the same dataset, which is described in greater detail in the next section.

The PRIMA survey contains a large amount of information on pupils in grades 2, 4, 6 and 8. For all pupils individual school performance is measured in the fields of arithmetic and language. These individual scores serve as the dependent variables which Bosker and Hox regress, in a multi-level framework, on a series of dummy variables measuring different class size brackets. The brackets are for the following class sizes: 5-9, 10-14, 15-19, 20-24, 25-29, 30-34 and 35-39. Although this specification captures some degree of possible non-linearity in the relation between class size and performance, the specification also ignores any variation within these brackets. The authors admit that their results may be biased because of endogeneity of class size, but no attempt is made to correct for this other than including a number of covariates. These include controls for the following: pupil’s gender and social background; teacher’s experience, gender and attitudes; and, class averages of pupils’ IQ, gender and social background.

The findings of Bosker and Hox provide a mixed picture. For grade 2 performance in arithmetic goes down when class size increases from 20-24 to 25-29; also, in classes in the size brackets 30-34 and 35-39 arithmetic scores are lower than in the reference category. It is important to note however, that increasing class size from 25-29 to 30-34 pupils actually increases performance. For the effect of class size on language performance in grade 2 a similar pattern emerges. Scores are significantly lower in the 25-29 and 35-39 size brackets, but not in the 30-34 category. Pupils in classes with 30-34 children perform as well on languages as pupils in classes with 5-9, 10-14, 15-19 or 20-24 children. This result is
especially important as 21% of the second grade classes in the sample fall in the 30-34 range, and another 27% in the 25-29 range.

Regressions pooling individuals in the 4th, 6th and 8th grades show that pupils in classes with more than 35 children (about 5% of the classes in the sample) do worse for both language and arithmetic. Adding 6-10 pupils to a class of 25-29 students causes a drop in average achievement by about 8 and 4 percentiles in language and arithmetic, respectively. Noticeable also is that for arithmetic pupils in the class size bracket 30-34 have the second highest scores.

These results do not support the hypothesis that, in general, smaller class sizes lead to better performance in the fields of arithmetic and language. Only for classes with over 35 pupils a negative effect of class size is found, but for classes with less than 35 children in it the pattern is erratic. The main shortcoming of the study is, however, that the results are not purged for possible endogeneity bias and, as the results in the study by Angrist and Lavy show, this may change the results dramatically. The current chapter assesses whether this is indeed the case.

1.3 Data and construction of variables
The following analysis makes use of PRIMA, a longitudinal survey containing information on Dutch pupils who were enrolled in grades 2, 4, 6 and 8 in the 1994/1995 school year. Several instruments were used in collecting the data: administrative sources, test results and questionnaires for the teachers, parents and headmasters of approximately 800 primary schools throughout the Netherlands. Of these, 400 were chosen to form a nationally representative sample of “regular” schools used in this analysis. Here, the sample is limited to only the observations on individuals from grades 4, 6 and 8, as the information on class size for grade 2 is too unreliable.

The PRIMA survey provides a wealth of information at the pupil, class, and school levels. We use pupils’ individual scores on arithmetic and language tests transformed into grade-level percentile rankings to measure school performance, and regress these scores on various pupil, class and school characteristics. At the individual level we include indicators of gender and weight factors accounting for socioeconomic status (SES) as covariates. The weight factor of a pupil indicates his/her social background and is used to determine the amount of

---

2 From the perspective of policy formulation it is interesting to mention that the researchers interpreted the same results rather differently. They state that “children in grade 2 who perform less well in arithmetic and language, do so partly because they were in classes with more than 25 children…” (p. 190). Obviously, this ignores the insignificant effect for languages in the 30-34 bracket and the rise in arithmetic score resulting from an increase in class size from the 25-29 to the 30-34 range. In our view, researchers who take the negative effects of larger classes on achievement seriously should pay equal attention to the positive effects of larger classes.

3 As mentioned, this data has previously been used in the study by Bosker and Hox (1996).

4 The documentation of the PRIMA survey states that a considerable number of teachers in grade 2 seem to have given erroneous information on class size. Probably they filled in the total number of pupils enrolled at that grade level instead of the number in their own class.
resources a school is given for a certain pupil; schools get more money for pupils with a disadvantaged background than for those with a "normal" background. The weight factor ranges from 1.00 to 1.90 and is defined as follows:

<table>
<thead>
<tr>
<th>Weight factor value</th>
<th>Definition of weight factor category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.90</td>
<td>Pupils with foreign-born parents that satisfy one of the following conditions: 1) father, mother or guardian has at most a VBO-level education; 2) the primary earning parent or guardian has either a job involving physical labor, or has no income from labor at the time of interview.</td>
</tr>
<tr>
<td>1.70</td>
<td>Pupils whose parents are transients.</td>
</tr>
<tr>
<td>1.40</td>
<td>Pupils living in a boarding school or a foster home, and whose parents are masters of a ship.</td>
</tr>
<tr>
<td>1.25</td>
<td>Pupils of whom either parents or guardians have at most an education at VBO-level.</td>
</tr>
<tr>
<td>1.00</td>
<td>Other pupils (reference group).</td>
</tr>
</tbody>
</table>

At the class level the regressions control for the following: class size, teacher’s gender and years of experience, class averages of pupil gender and weight factor, dual teacher class (whether more than one teacher has taught the class), and multi-grade class (whether the class combines individuals from more than one grade level). School level variables include three dummy indicators denoting the denomination of the school, the average SES within a school, and total enrollment of the school. The sample has been limited to only those observations without missing values for the variables used in the analysis.

The dataset also contains a measure of pupils’ IQ. This IQ score is based on tests taken at the same time as those that measure language and arithmetic scores. Consequently, this IQ measure can not be regarded as an independent measure of innate ability and is therefore not included as a control variable in the regressions. In our analysis we use, however, the variable “number of classmates with similar IQ” which is constructed on the basis of pupils’ IQ scores. This variable will be used to identify a peer group effect motivated by social cognitive learning theory (see the next section).

As we want to investigate if pupils in smaller classes perform better than those in larger classes, our most crucial explanatory variable is class size. Because of the possible endogeneity of class size, OLS estimates of the class-size effect on pupils’ test scores may be biased. We therefore take an IV approach, using an instrument for class size based on the Dutch rules that link total enrollment in the school to the number of teachers a school receives funding for. These rules are decreed by the Dutch Ministry of Education, Culture and Science in the

---

5 VBO stands for Voorbereidend Beroepsonderwijs or secondary vocational education, which serves as the lowest form of secondary schooling in the Dutch educational system.

6 Note that all independent variables except for those based on IQ measures denote lagged indicators of individual, class and school characteristics the year prior to the survey (1993-94), whereas achievement measures are proxied by scores on tests taken within the first few months of the survey year (1994-95).

7 The three categories of religion/ideology are Protestant, Roman Catholic and “Other” which include Islamic, general special and other ideologies. Public schools serve as the reference group.

8 The total number of observations contained in the PRIMA reference sample are 5,698, 5,368 and 5,608 for the 4th, 6th and 8th grades, respectively. The number dropped due to missing values in the baseline variables are 965, 1,007 and 1,018 leaving a total of 4,733, 4,361 and 4,590 usable observations, respectively, for 4th, 6th and 8th grades.

9 Note that Angrist and Lavy’s instrument is based on enrollment at the grade level, whereas our instrument is based on total enrollment at the school level.
“Formatiebesluit WBO”. In order to determine the number of teachers a school receives funding for, first the weighted number of students (WNS) of a school is calculated using the following formula:

$$WNS = \text{INT}(1.03 \times \text{max} \{\text{INT}(\sum_{i=1}^{N} w_i - 0.09 \times N), N\})$$

(1)

where $N$ is total enrollment in the beginning of a school year and $w_i$ is the weight factor of pupil $i$ as described above. The weighted sum of all students is reduced by 9% of the total number of students (as it is assumed that the school should be able to cope with this proportion of disadvantaged students without additional resources), and rounded down to the nearest integer. If this number is lower than the actual (unweighted) number of students then the actual number of students $N$ is used instead. The outcome is increased by 3% to account for the possible increase in total enrollment during the school year due to new entrants. The number of teachers that can be appointed given the weighted number of students can then be looked up in a table. The schedule shows that schools with a number of weighted students up to 30 can appoint 2 full-time teachers, and the number of teachers increases by 0.2 or 0.3 full-time units for every 6 to 11 additional weighted students. Figure 1.1 shows that the resulting pupil-teacher ratio function is discontinuous; the pupil-teacher ratio increases to 15 when the weighted number of students goes up to 30, it then drops to about 14, climbs up again to more than 16, drops to 15.4, etc. From the characteristic “saw tooth” shape of the figure the function is clearly non-monotonic and discontinuous and thus, may serve as an instrument for class size.

Figure 1.1 - Graph of enrollment function determined pupil-teacher ratio
When comparing the graph of our instrument with that of Angrist and Lavy's, two differences are apparent. Firstly, the differences at the points of discontinuity are much larger in their case than in ours. The distance from a peak to a trough represents a change in average class size by 1 or 2 with our instrument, while using Maimonides' rule it changes abruptly from 40 to 20.5. While it seems to be more attractive to have large exogenous variations to identify the effect of class size reduction, it is important to stress that the above-mentioned Dutch policy proposed long run class size reductions of 3 to 4 pupils. In that case small reductions of 1 or 2 is more on the mark than reductions exceeding 15 pupils. Secondly, the number of discontinuities is much smaller in the Israeli than in the Dutch case. With a small number of discontinuities only the observations near one of these (discontinuous) points are relevant for the identification of the class size effect. Consequently, Angrist and Lavy present separate results for those observations that are in a -5/+5 range of the points of discontinuity (40, 80, 120 pupils per grade level). With so many evenly spread points of discontinuity, as implied by the Dutch funding scheme, all observations are in a relevant range of potential treatment. Hence, we do not have to narrow our sample. To put it differently, while Angrist and Lavy identify Local Average Treatment Effects, the nature of the instrument used here allows identification of the Average Treatment Effect.

For each school we know the number of pupils with various weight factors, and so we can derive the number of group teachers that can be appointed by the school based on the ministerial rules. The instrument we use in our estimations is predicted average class size, calculated as total school enrollment divided by the predicted number of group teachers using the ministerial rules. Henceforth, we refer to this instrument as the "Enrollment Function Determined" class size or EFD class size for short. A nice feature of this instrument is that it is a discontinuous function of the school's average SES weight factor and total enrollment. This allows us to include both the school's average SES weight factor and total enrollment as regressors in the second stage equation. Consequently, it is unlikely that the instrument is picking up any direct effects of these two variables. Hence, for our instrument to be valid it is only necessary that the particular transformation of average SES weight and total enrollment affect achievement only through its effect on class size, controlling for linear average SES weight and total enrollment.\footnote{In Section 1.5.2 we report results which show that including higher order terms of total enrollment and school average SES weight do not affect our findings.}

Appendix 1A contains descriptive statistics by grade level for the variables used in the analysis. A brief look at the table reveals nothing too surprising. Class sizes are remarkably similar across grades; the average 4th grade class has between 25 and 26 students and increases to between 26 and 27 for grades 6 and 8, respectively. Over half of all students are not considered disadvantaged as measured by the SES weighting scheme (i.e. their SES weight factor equals 1.00). Another 35 to 40% are slightly disadvantaged with an SES weight of 1.25. What is perhaps most noteworthy concerns teacher characteristics. In particular, the proportion of instructors that are female drops sharply as grade level increases; whereas over three quarters of the 4th grade teachers are women, this fraction drops to under one half in 6th grade and just over one quarter in 8th grade. In addition, the average 8th grade instructor has over one year more experience than their counterparts teaching grades 4 and 6.
1.4 Empirical analysis

For the analysis of the effect of class size on pupils’ performance in Dutch primary schools, we estimate equations of the form

\[
y_{ics} = \alpha + X_i' \beta + C_c' \gamma + S_s' \delta + n_{ics} \eta + \varepsilon_{ics} \tag{2}
\]

\[
n_{ics} = \theta + X_i' \kappa + C_c' \lambda + S_s' \mu + efd_s' \pi + \nu_{ics} \tag{3}
\]

where \(y_{ics}\) is the percentile score (in arithmetic or language) of pupil \(i\) in class \(c\) of school \(s\); \(X_i\), \(C_c\) and \(S_s\) are matrices containing variables pertaining to the individual, class and school, respectively; \(n_{ics}\) is the number of pupils in pupil \(i\)’s class \(c\) within school \(s\); \(efd_s\) is the EFD class size of school \(s\); and \(\varepsilon_{ics}\) and \(\nu_{ics}\) are random components of the error term specific to the pupil. As pupils within the same school may have interdependent error terms, equation (2) is calculated with robust standard errors which allow for correlated error terms of pupils within the same school.\(^{11,12}\) The difference between the equations estimated in this chapter and the ones estimated by Angrist and Lavy is that here we take individual pupils as the unit of analysis while in the Angrist and Lavy paper the unit of observation is the class.\(^{13}\)

Our main findings are comprised only from the results using actual class size as reported by the teacher. In the next section we explore whether there are differences when alternative measures such as pupil-teacher ratio and average class size (calculated as administrative reported enrollment divided by number of teachers) are implemented. In addition, results for other specifications and different sub-samples are reported. We performed all analyses separately for grades 4, 6 and 8. The numbers of observations are large enough to allow that, and we do not want to impose any a priori restrictions on the effects of the various regressors on school performance at different grade levels.\(^{14}\)

The three parts of Table 1 contain our main results. For each grade level five columns with results are reported. The first columns report the first-stage estimation results. The second and fourth columns give the respective OLS (baseline) estimates for arithmetic and language achievement while columns three and five contain the corresponding IV results. We do not report the full estimation results, but report only the coefficients that are of prime interest from the perspective of the causal effect of class size on achievement. Hence, we do not report the results for all of the covariates, but only those for total enrollment and the average SES weight. These are reported because our instrumental variable is a direct transformation of the two. Included in each cell of the tables is the estimated coefficient with the robust standard errors corrected for grouped (at the school level) heteroskedasticity in parentheses.

\(^{11}\) Robust standard errors are calculated using Roger’s (1993) generalization of the estimator put forth by Huber (1967), which takes into account possible correlations between the errors of observations within groups (in our context schools). The exact procedure we use for obtaining robust variance in the presence of a group clustered error structure can be found in StataCorp (2001), pp. 254-258.

\(^{12}\) It is plausible that individual error terms may also be correlated within classes. However, for reasons related to our identification strategy, accounting for grouped correlation at the school level is more proper (a more in-depth discussion follows below in Chapter 3, Section 3.5.2.3).

\(^{13}\) We checked for our data whether raising the unit of analysis leads to different conclusions; this is not the case.

\(^{14}\) Tests of differences in class size effects on achievement between grades reveal significant differences using both OLS and IV for language and using IV for arithmetic.
Starting with the 4th grade pupils, the reduced form (first-stage) results in the first column show that EFD class size is highly correlated with the potentially endogenous actual class size at grade 4. The partial correlation coefficient equals 0.38. The OLS results in columns two and four show an insignificant positive effect of class size on pupils’ arithmetic percentile scores and an insignificant negative effect on language percentile scores, respectively. These are results from an equation including a large number of control variables. Without control variables, the class size effects in both equations are significantly positive. Although the IV point estimates in columns three and five differ largely from the OLS results, the differences are not significant as indicated by Hausman tests, which accept the null hypothesis that any endogeneity of class size has a negligible effect on scholastic achievement. For both fields of study (arithmetic and language) the IV procedure yields increased but nevertheless insignificant point estimates of the class size effect. Hence, for pupils in grade 4 the estimation results that correct for endogeneity bias give no support for the view that pupils benefit from being placed into smaller classes.

The middle part of Table 1 reports results for 6th graders. EFD class size is again highly correlated with actual class size (with a partial correlation now equal to 0.47). The OLS and IV results for percentile scores in arithmetic are similar to those obtained for 4th graders. The estimates of the class size effect are not significantly different from zero, and going from the OLS to IV result turns the point estimate from negative into positive. For languages, the results are somewhat different; the OLS-estimate is significantly negative (at the 10%-level) indicating that a reduction of class size is beneficial for language achievement. The IV-estimate suggests, however, that this inference might be attributed to endogeneity bias.

Lastly, results for 8th grade pupils are reported in the left portion of the table. Again EFD class size is highly correlated with the potentially endogenous variable actual class size; the partial correlation equals 0.43. Notice that for 8th graders the first-stage equation also includes significantly positive coefficients for total enrollment and school average SES weight, while for 4th and 6th graders this was not the case. The significance of total enrollment and school average SES weight do not, however, result in a weaker instrumental variable. The OLS and IV results for 8th graders also differ from those obtained for grades 4 and 6. The OLS estimates of the class size effect do not differ significantly from zero, but the IV estimates do. For both arithmetic and language, we find that 8th graders benefit from being placed in larger classes. For both IV estimations, the Hausman test rejects the null hypothesis that any possible endogeneity in the class size variable has a negligible effect on scholastic achievement. Thus, using an IV procedure is warranted.

The fact that endogeneity is only a problem for 8th grade and not for 4th and 6th grades suggests that the richness of the dataset accounts for much of the potential endogeneity in the class size variable. As a result the OLS estimates for these grade levels are not significantly different from the IV estimates. This in turn supports the approach followed by Dearden et al (2000) who identify the causal effect of school quality on educational attainment and wages by using a very extensive and rich set of control variables.
### Table 1 - Effect of teacher reported class size on arithmetic and language achievement of 4th, 6th and 8th graders

<table>
<thead>
<tr>
<th></th>
<th>4th grade</th>
<th></th>
<th></th>
<th>6th grade</th>
<th></th>
<th></th>
<th>8th grade</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First stage</td>
<td></td>
<td></td>
<td>First stage</td>
<td></td>
<td></td>
<td>First stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arithmetic</td>
<td>OLS IV</td>
<td>OLS IV</td>
<td>Arithmetic</td>
<td>OLS IV</td>
<td>OLS IV</td>
<td>Arithmetic</td>
<td>OLS IV</td>
<td>OLS IV</td>
</tr>
<tr>
<td><strong>EFD class size</strong></td>
<td>1.322*** (0.179)</td>
<td>0.093</td>
<td>0.137</td>
<td>0.091</td>
<td>0.223</td>
<td>0.207</td>
<td>0.261</td>
<td>0.064</td>
<td>0.857</td>
</tr>
<tr>
<td><strong>Class size</strong></td>
<td>0.145 (0.147)</td>
<td>0.123</td>
<td>0.371</td>
<td>0.136</td>
<td>0.273</td>
<td>0.120*</td>
<td>0.270</td>
<td>0.131</td>
<td>0.367***</td>
</tr>
<tr>
<td><strong>Total enrollment</strong></td>
<td>0.007*** (0.005)</td>
<td>0.014</td>
<td>0.010</td>
<td>0.005</td>
<td>0.009</td>
<td>0.013</td>
<td>0.014</td>
<td>0.007</td>
<td>0.012</td>
</tr>
<tr>
<td><strong>Adjusted R²</strong></td>
<td>0.3299</td>
<td>0.0842</td>
<td>0.0828</td>
<td>0.1420</td>
<td>0.1402</td>
<td>0.4257</td>
<td>0.0816</td>
<td>0.0787</td>
<td>0.1186</td>
</tr>
<tr>
<td><strong>Partial R²</strong></td>
<td>0.3839***</td>
<td>0.467***</td>
<td>0.4310***</td>
<td>0.467***</td>
<td>0.4310***</td>
<td>0.467***</td>
<td>0.4310***</td>
<td>0.467***</td>
<td>0.4310***</td>
</tr>
</tbody>
</table>


Dependent variable in first stage is actual class size as reported by the teacher. Dependent variable in second-stage equations is pupils’ percentile rankings on standardized arithmetic and language tests.

All regressions include controls for the following: four individual SES-class dummies, dummy for pupil’s gender, percentage of girls in pupil’s class, teacher’s gender and experience, dummy for dual teacher class, dummy for multi-grade class, three dummy variables for school’s denomination, and a constant.

Robust standard errors taking account of correlated disturbance terms within schools are reported in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%.
The results reported in Table 1 for different grade levels reveal a remarkably consistent pattern. The point estimates of the class size effect are almost never significantly negative and in some cases they are even significantly positive. Krueger (1999) reports an improvement of 5-7 percentile points from being placed in a small (13-17) instead of a large (22-25) class. Thus, the class size effect equals a 0.7 percentile point gain per one pupil reduction. When we take this class size effect - which is arguably the most convincing result in the class size literature - as the null hypothesis, we reject it in 11 out of 12 cases.

Our findings indicate that Dutch primary school pupils do not benefit from being placed in smaller classes. As the results reported in Section 5 below show, the findings are very robust. Including higher order terms of total enrollment and school average SES weight, using different measures of class size, allowing for nonlinearities or limiting the sample to specific subgroups does not affect the main results.

The findings reported here not only contradict the "common wisdom" dominant among educationalists but are also at odds with the results from other recent studies using the IV method (but are in line with Hoxby's results). Although the results should be taken seriously as they stand, the findings will gain importance when their underlying mechanism can be unveiled. It is to this issue that we now turn.

When pupils in larger classes learn no less than pupils in smaller classes, an explanation might have something to do with children learning from one another. More specifically, when there is a larger number of pupils in a class there are potentially more children from whose presence a pupil may benefit. Social cognitive learning theory strongly suggests that the achievement of pupils may benefit from the presence of similar classmates. This theory emphasizes that people acquire knowledge, skills and attitudes by observing others (Bandura, 1986; Schunk, 1991). Moreover, educational practices like the use of peer models and ability grouping of pupils are based on the idea that pupils learn by observing and working together with other similar individuals. In this literature, evidence can be found of positive effects of both ability grouping (Slavin, 1987) and peer modeling (Schunk, 1987) on the school performance of pupils.

Research on peer modeling suggests that both the competencies and number of role models can affect pupils' performances. Similarity in competence (which we proxy with IQ) appears important for purposes of self-evaluation. France-Kaatrude and Smith (1985) show that children who were allowed to compare their performances with a similarly performing peer compared themselves more often, and demonstrated greater task persistence than children who were offered comparisons with superior or inferior peers. Observing the success of individuals similar to oneself raises the observer's self-efficacy and motivates him or her to perform the same tasks. This is especially true among children who have experienced difficulties in learning skills (Schunk and Hanson, 1985). Moreover, Schunk et al. (1987) and Bandura and Menlove (1968) find advantages for multiple peers, presumably as these increase the probability that observers will perceive themselves as similar in competence to at least one
of the role models. Especially when subjects doubt their learning capabilities, observation of multiple peers succeeding may better promote subjects' self-efficacy (Bandura, 1986).15,16

Given the pupil's own IQ, the expected number of other pupils in the class with similar IQ increases with class size.17 Whether this mechanism indeed explains a zero net effect of class size reduction can be tested by including the number of pupils in the class with similar IQ as an additional regressor in the test score equations. As the PRIMA data include for each pupil in the sample his/her own IQ as well as the IQ score of each of its classmates, we are able to construct the variable “number of classmates with similar IQ”. This variable has an expected value that increases with the size of the class.18 Our hypothesis is thus twofold: inclusion of this variable should reduce (or possibly even reverse) positive class size effects on achievement, and the new variable should have a positive direct effect on achievement. To make this test operational we have defined students with similar competence as those in a class that fall within a range of plus or minus half a grade-level standard deviation (+/- 2 IQ-points) around the given pupil's IQ.19 The correlation coefficient between this variable and actual class size equals 0.46 in grade 4, 0.31 in grade 6, and 0.40 in grade 8. It should also be noted that the number of similar pupils is not a direct measure of class homogeneity. Class homogeneity is best measured by a statistic such as variance. Although the number of pupils in the class with similar IQ and the variance of IQ in a class are negatively correlated, this correlation is weak: -0.25 in grade 4, -0.18 in grade 6, and -0.16 in grade 8.

Table 2 presents estimation results from specifications identical to those reported in Table 1, except that the additional control variable “number of classmates with similar IQ” is included.20 In all but one case does the newly created variable have the predicted positive effect on performance and prove to be highly significant. For grade 4, we find a drop in the class size effects on arithmetic and language achievement for both the baseline OLS estimates as well as those produced by the IV procedure. Most notably, for language achievement the negative OLS baseline estimate proves to be significant at the 10%-level. The results for grade 6 are even stronger: all point estimates of the class size effect drop considerably, and

15 Another possible mechanism leading to a positive relation between performance and the number of similar classmates is that teachers may target their instruction to groups of children with similar competence. If instruction time is divided in proportion to the size of the group, members of larger groups receive more instruction time.

16 In a recent paper Lazear (2001) proposes another peer group mechanism in relation with class size effects. The driving force in his model is the presence of disruptive children who cause negative externalities (congestion). To explain our findings, we are, however, in search for a mechanism causing positive externalities.

17 An alternative hypothesis is that children learn from individuals that are smarter than themselves. We tested this hypothesis by including number of students in an individual's class with an IQ higher than their own. Even after controlling for pupils' own IQ, this served merely as a proxy for IQ rank within a class producing a large negative effect with no significant decrease in the positive class size effect.

18 As mentioned earlier, IQ was measured at the same time as the test scores. Although individual IQ scores have probably been affected by the size of the class, there seems no obvious reason to assume that class size affects the number of classmates with similar IQ other than through the simple mechanism by which the expected number of similar classmates increases with class size. Notice that with IQ measured after class assignment, explicit ability tracking on the basis of measured IQ is prevented.

19 Actually our definition is conservative as 2 IQ points is a bit smaller than one half of an IQ standard deviation for all three grades (see descriptive statistics in Appendix IA).

20 We also estimated the same models including the class average IQ as an additional covariate. The results are similar to the ones reported here.
three of the four are significantly negative. In all cases the number of classmates with similar IQ has the predicted positive effect on achievement. The results for grade 8 also support the hypothesis set out above; the number of classmates with similar IQ has the predicted positive effect on achievement. The OLS estimates of the class size effect change from positive to negative after including the number of similar classmates as additional regressor. The two IV estimates of the class size effect for grade 8 that were significantly positive without a control for number of classmates with similar IQ have decreased in magnitude and significance. The IV class estimate for arithmetic achievement is only marginally significant while that of language does not differ from zero at any reasonable level of significance.

In conclusion, the non-negative, and sometimes even positive, relation between class-size and school performance in Dutch primary schools can - at least partially - be attributed to the fact that reduction of class size also reduces the (expected) number of pupils in the class with a similar level of competence. This reduction apparently limits a pupil’s scope to learn from her or his classmates.

1.5 Robustness of results
This section addresses the robustness of the main results by experimenting with different measures of class size, including higher order terms of total enrollment and school average SES weight as covariates, imposing a non-linear functional form of class size, and using limited sub-samples.21

1.5.1 Alternative measures of class size
As alternatives to the actual class size as reported by the teacher, the pupil-teacher ratio for each class has been constructed in addition to average class size within a school. The results for these two alternative class size measures are remarkably similar to those reported above in our main findings. In all but a few cases both the sign and significance level of the point estimates are the same. Therefore, we conclude that the findings are robust to differing measures of class size.

1.5.2 Inclusion of higher order terms of total enrollment and school average SES weight
The instrumental variable EFD class size plays a key role in determining the causal effect of class size on achievement. For this variable to be a valid instrument it must have no other effect on achievement other than through its effect on class size. As the instrumental variable is derived from official ministerial tables which translates SES weighted total enrollment into

21 To save space, we do not present tables with results of the different specifications. These are available from the author upon request.
<table>
<thead>
<tr>
<th></th>
<th>4th grade</th>
<th></th>
<th>6th grade</th>
<th></th>
<th>8th grade</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Arithmetic</td>
<td>Language</td>
<td>Arithmetic</td>
<td>Language</td>
<td>Arithmetic</td>
<td>Language</td>
</tr>
<tr>
<td>EFD class size</td>
<td>1.176</td>
<td>(0.155)**</td>
<td>1.277</td>
<td>(0.184)**</td>
<td>1.200</td>
<td>(0.154)**</td>
</tr>
<tr>
<td>Class size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.139</td>
<td>0.084</td>
<td>-0.239</td>
<td>0.020</td>
<td>-0.261</td>
<td>-0.033</td>
<td>-0.340</td>
</tr>
<tr>
<td>(0.151)</td>
<td>(0.411)</td>
<td>(0.135)*</td>
<td>(0.435)</td>
<td>(0.138)*</td>
<td>(0.280)</td>
<td>(0.120)*</td>
</tr>
<tr>
<td>Number of students</td>
<td>0.407</td>
<td>1.074</td>
<td>0.969</td>
<td>0.552</td>
<td>0.430</td>
<td>0.203</td>
</tr>
<tr>
<td>with similar IQ</td>
<td>(0.057)**</td>
<td>(0.153)**</td>
<td>(0.240)**</td>
<td>(0.225)**</td>
<td>(0.316)</td>
<td>(0.055)**</td>
</tr>
<tr>
<td>Total enrollment</td>
<td>0.004</td>
<td>0.010</td>
<td>0.008</td>
<td>0.010</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>(0.004)</td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.011)</td>
<td>(0.013)</td>
<td>(0.005)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.4013</td>
<td>0.1031</td>
<td>0.1016</td>
<td>0.1468</td>
<td>0.1448</td>
<td>0.4416</td>
</tr>
<tr>
<td>Partial R^2</td>
<td>0.3612***</td>
<td></td>
<td></td>
<td>0.4467***</td>
<td></td>
<td>0.4070***</td>
</tr>
<tr>
<td>Hausman test</td>
<td>-0.256</td>
<td>-0.298</td>
<td>-0.284</td>
<td>0.186</td>
<td>-0.294</td>
<td>-0.284</td>
</tr>
<tr>
<td>(0.454)</td>
<td>(0.473)</td>
<td>(0.305)</td>
<td>(0.340)</td>
<td>(0.305)</td>
<td>(0.340)</td>
<td>(0.305)</td>
</tr>
</tbody>
</table>

Dependent variable in first stage is actual class size as reported by the teacher. Dependent variable in second-stage equations is pupils' percentile rankings on standardized arithmetic and language tests. All regressions include controls for the following: four individual SES-class dummies, dummy for pupil's gender, percentage of girls in pupil's class, teacher's gender and experience, dummy for dual teacher class, dummy for multi-grade class, three dummy variables for school's denomination, and a constant. Robust standard errors taking account of correlated disturbance terms within schools are reported in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%.
the number of class teachers, it is important that the instrument does not pick up any direct
effects from a school’s SES weighted total enrollment on achievement. In Section 4 we
controlled for this by including the school’s average SES weight and total enrollment as
covariates, but perhaps it is insufficient to enter these terms in linear form only.

To this end, we estimated specifications including quadratic terms in total enrollment
and average SES weight as well as an interaction term between total enrollment and average
SES weight. These results are fairly similar to those in Tables 1 and 2. For the estimations
without “number of classmates with similar IQ”, the main difference is that the IV estimates
of the class size effects for 8th graders are not significantly positive, while they are so in Table
1. For the specification including the "number of classmates with similar IQ", inclusion of the
higher order terms of total enrollment and school average SES weight cause a drop of the
coefficient of EFD class size in the first-stage equation. Also, although the partial correlation
coefficient is lower, we still end up with a fairly strong instrument. The terms in enrollment
and school average SES weight do in most cases not have a significant impact on school
performance.\textsuperscript{22}

1.5.3 Non-linear class size

The above findings implicitly assume that the relationship between class size and scholastic
achievement is of a linear form, which may not be the case. To address the possibility of
nonlinearities, we changed the functional form of our model by using a two-step selection
model with class size measured in 7 different categories. Following Bosker and Hox (1996),
we broke class size into the following categories: 5 to 9; 10-14; 15-19; 20-24; 25-29; 30-35;
35 and above. In the first step we treated class size as an ordinal variable, and applied an
ordered probit equation, in the second step the percentile achievement scores were just as
before treated as cardinal variables. The identifying variable in this approach is the instrument
EFD class size.

When “number of classmates with similar IQ” is omitted, the results for 4th grade math
achievement suggest that relative to classes with 25 to 30 students, those with more than 35
have a significant positive effect, while being placed in small classes with 5 to 9 pupils has a
significant negative impact on achievement. The selectivity-correction term is insignificant
indicating that the endogeneity of class size has no effect on the OLS point estimates. The
only significant effect of class size on 4th grade language achievement is for individuals in
class sizes with 20 to 24 students (one category lower than the reference) who are estimated to
score approximately 4 percentile points higher. For 6th graders, the relation between class size
categories and achievement appears to be completely flat, with the exception of the negative
effect of being placed in a very small class (5-9) for the arithmetic score. The estimates for
the 8th grade make for a completely different story. Here, significant effects on math and
language achievement of negative 3 to 4 percentile points are associated with the downsizing
of a class from the range of 25-29 to 20-24 students.

With “number of classmates with similar IQ” included, the results confirm those
obtained before. The significantly negative effect on arithmetic scores of being placed in a

\textsuperscript{22} This remains the case even after further smoothing the achievement/enrollment relation by including a cubic
term in enrollment or a series of dummy variables for different total enrollment categories.
small 4th grade class vanishes, as does the significantly positive effect of being placed in a very large class. The same holds (in most cases) for the effects on arithmetic and language achievement of being in an 8th grade class of 20-24 students.

The class size effects presented in Table 1 point to a flat achievement/class size profile. The results from the non-linear specification suggest that - with a few exceptions - the profile is flat everywhere. Formal tests confirm this. Not only are most of the class size coefficients not significantly different from zero, they are also not significantly different from each other. Of the 90 possible pairwise comparisons of IV estimates, only 13 point to a significant difference.

1.5.4 Homogeneity with respect to SES
One possible explanation as to why we do not find the commonly expected negative correlation between class size and achievement may lie in composition of class with respect to SES. Indeed, policy made by the Ministry of Education encourages a higher allocation of resources for those with lower SES. It would then be no surprise that those students with lower SES end up in smaller classes and have lower achievement on average, driving the result of a positive class size/achievement relationship. To test for this, we restrict our sample by preserving observations that are in relatively homogeneous classes with respect to SES. To be more precise, we limit the observations only to those unweighted individuals (SES weight equals unity) that are in classes with an average SES weight of 1.35 or less.

For the specification without “number of classmates with similar IQ”, results are almost identical to those in Table 1. For the specification with “number of classmates with similar IQ” the results differ somewhat from those in Table 2. None of the class size effects are significantly negative. Also the point estimates of the effect of the number of classmates with similar IQ are somewhat smaller (but remain highly significant in all but four cases). This suggests that pupils from a more disadvantaged social background benefit more from both the “pure” class size effect, and the presence of classmates with similar IQ.

1.5.5 Schools with one class per grade level
Endogeneity bias in the estimated effect of class size on achievement may arise from two different sources. One is school selection by the parents; the other is sorting of pupils over classes by the schools. An easy remedy against biases from this second source is to restrict the sample to observations that have only one class at the grade level. As a result of the small scale of Dutch primary schools, selecting on the basis of this criterion leaves us with 80 percent of the observations. Repeating the analysis for this 80 percent generates results that are almost identical to those reported in Tables 1 and 2.

1.6 Concluding remarks
This chapter analyses the causal effect of class size on scholastic performance in Dutch primary schools. Empirical estimation of this effect is difficult due to the possible endogeneity of class size. Two sources of endogeneity may play a role: choices by parents, where parents who care more about their children’s achievement opt to send their children to
schools with smaller classes, and selection by schools where weaker pupils are assigned to smaller classes. We correct for these biases via an instrumental variables approach where the instrument is based on ministerial rules determining the number of class teachers as a function of enrollment. The resulting class size function is discontinuous and thus, can be used to isolate the exogenous portion of class size. The estimation results show that after correcting for endogeneity, pupils in large classes do no worse - and sometimes even better - than identical pupils in small classes. This result holds for a large variety of functional forms, specifications and using restricted samples.

As an explanation for this result, we formulate a hypothesis based on the social cognitive learning literature where pupils’ achievement benefits from a larger number of classmates with similar levels of competence. By reducing class size, the (expected) number of classmates with similar IQ falls, which may have a negative effect on achievement. We test this hypothesis by constructing a new variable that counts for each individual the number of pupils in the class with a similar IQ. In all specifications this new variable has the predicted positive effect on performance and is in most cases highly significant. Moreover, of the twelve class size effects that we estimate: two of the significantly positive estimates drop to levels of marginal significance and insignificance, respectively, when the number of similar classmates is included; two estimated effects remain insignificantly positive; four effects change from insignificantly positive to insignificantly negative; three effects change from insignificantly negative to significantly negative; and one significantly negative effect becomes highly significant.

It is important to realize that the new variable “number of classmates with similar IQ” differs from conventional measures of class homogeneity. Class homogeneity is measured in terms of the dispersion of IQ and is not intended to be a measure varying with the number of pupils in the class. Instead, our measure explicitly intends to do so in order to provide a mechanism that can explain the non-negative effect of class size on scholastic performance.

The important policy conclusion to be drawn from the findings in this chapter is the following: class size reduction at the scale currently implemented in the Netherlands is unlikely to be an effective intervention to boost performance. The reason is that reduction of class size does not only decrease the pupil-teacher ratio, but also decreases the (expected) number of classmates with a similar level of competence. It is tempting to draw policy conclusions based on the finding that a larger number of classmates with similar IQ increases achievement. This finding may even be regarded as a plea for ability tracking. It is important to realize, however, that the two achievement measures used in this chapter (arithmetic and language) do not cover the entire range of skills and attitudes learned in schools. Schools also play a role in developing good citizenship which includes (in our view) interacting with others who are dissimilar in a diverse environment.

Perhaps the best way to summarize the findings of this chapter is to consider the following situation. A child comes home and tells her parents that the principal, in an effort to reduce class sizes, has decided to take three children out of her class. Should the parents of this child be happy with the message? According to our results, this depends on which children are taken out of the class. If these are children who stimulate the child’s learning behavior, the parents should not favor this class size reduction.
### Appendix 1A  Descriptive statistics of selected variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Grade 4</th>
<th>Grade 6</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
<td>Minimum</td>
</tr>
<tr>
<td>Math score (absolute)</td>
<td>1042.08</td>
<td>67.67</td>
<td>822.70</td>
</tr>
<tr>
<td>Math score (percentile)</td>
<td>52.33</td>
<td>27.67</td>
<td>1.00</td>
</tr>
<tr>
<td>Language score (absolute)</td>
<td>1030.68</td>
<td>36.54</td>
<td>841.80</td>
</tr>
<tr>
<td>Language score (percentile)</td>
<td>54.75</td>
<td>27.77</td>
<td>1.00</td>
</tr>
<tr>
<td>Class size</td>
<td>25.47</td>
<td>5.76</td>
<td>5.00</td>
</tr>
<tr>
<td>Number of students with similar IQ</td>
<td>6.03</td>
<td>4.34</td>
<td>0.00</td>
</tr>
<tr>
<td>School enrollment</td>
<td>216.22</td>
<td>103.08</td>
<td>23.00</td>
</tr>
<tr>
<td>Weight factor = 1.00</td>
<td>0.54</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Weight factor = 1.25</td>
<td>0.35</td>
<td>0.48</td>
<td>0.00</td>
</tr>
<tr>
<td>Weight factor = 1.40</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>Weight factor = 1.70</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>Weight factor = 1.90</td>
<td>0.10</td>
<td>0.30</td>
<td>0.00</td>
</tr>
<tr>
<td>Gender of student (1-Female)</td>
<td>0.49</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Percent of class female</td>
<td>0.49</td>
<td>0.12</td>
<td>0.00</td>
</tr>
<tr>
<td>Class average weight factor</td>
<td>1.18</td>
<td>0.17</td>
<td>1.00</td>
</tr>
<tr>
<td>Teacher experience</td>
<td>17.59</td>
<td>8.15</td>
<td>1.00</td>
</tr>
<tr>
<td>Dual teacher class</td>
<td>0.38</td>
<td>0.49</td>
<td>0.00</td>
</tr>
<tr>
<td>Gender of teacher (1-Female)</td>
<td>0.84</td>
<td>0.37</td>
<td>0.00</td>
</tr>
<tr>
<td>Multi-grade class</td>
<td>0.27</td>
<td>0.44</td>
<td>0.00</td>
</tr>
<tr>
<td>Public</td>
<td>0.27</td>
<td>0.45</td>
<td>0.00</td>
</tr>
<tr>
<td>Protestant</td>
<td>0.22</td>
<td>0.41</td>
<td>0.00</td>
</tr>
<tr>
<td>Roman Catholic</td>
<td>0.44</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Other</td>
<td>0.07</td>
<td>0.25</td>
<td>0.00</td>
</tr>
<tr>
<td>School average weight factor</td>
<td>1.18</td>
<td>0.17</td>
<td>1.00</td>
</tr>
<tr>
<td>IQ</td>
<td>26.18</td>
<td>5.39</td>
<td>3.00</td>
</tr>
</tbody>
</table>

Number of observations: 4,733, 4,361, 4,590

* Number of observations for EFD class size equals 4,453, 4,090 and 4,244 for the 4th, 6th and 8th grade, respectively.