ESTIMATING THE EFFECT OF CLASS SIZE ON STUDENT ACHIEVEMENT: THE CASE OF HUNGARY

By
Gábor Kiss

Submitted to
Central European University
Department of Economics

In partial fulfilment of the requirements for the degree of Master of Arts in Economics

Supervisor: Professor Gábor Kőrösi

Budapest, Hungary
2015

Abstract

This thesis seeks to estimate the effect of class size on the achievement of students on the National Assessment of Basic Competences, a yearly administered standardized test in mathematics and reading comprehension. The analysis is based on the 6th and 8th grade test scores of the cohort of students who were 6th graders in 2011. The difficulties of the estimation stem from the endogenous relationship between enrollment and achievement: students enroll in classes based on their observable and unobservable characteristics and the interplay of the objectives of parents and the school management. I attempt to measure the effects by utilizing the panel nature of the dataset and two sources of supposedly exogenous variation in class size: 1) the maximum class size rule in force, and 2) the closure of schools in the school year of 2011/2012, and the fact that the re-enrollment of students in other schools is also regulated by the law on public education. I argue that none of these methods rule out the endogeneity issue. The 2SLS estimates do not yield statistically significant coefficient estimates, while the first differenced regressions show small negative effects.

Keywords: education, school, skill, instrumental variables, panel data

Acknowledgements

I am most grateful to my supervisor, Professor Gábor Kőrösi for the valuable discussions and comments he provided during the work on this thesis.

I would also like to thank Tamara Stalzer and my family for their support. Finally, I would like to thank Bálint Herczeg and Klára Major for helping me with Stata codes and the data.

Table of Contents

Introduction

Class size reduction has been a popular policy tool in the United States aimed at increasing student achievement. Several US states including California, Texas and Florida have implemented class-size caps in the past few decades partly as a result of empirical findings (Schanzenbach, 2014). This thesis sets out to examine the effect of class size on the test scores of students, a subject that has a relatively broad literature not only in the field of education research, but in economics as well. The empirical investigation is based on a Hungarian panel dataset with student and school level data – the database of the National Assessment of Basic Competences (NABC). This enabled me to use some of the more common methods in the literature widely applied to study the relationship between class size and student achievement, and also to employ a somewhat different approach relying on the panel nature of the dataset.

Although achievement on standardized tests is usually not the direct objective of human capital investment activity, they may "matter" in multiple ways: if tests measure skills that are truly useful on the job market, increasing these skills and producing higher test scores directly results in higher lifetime income on average. The test scores and the skills they measure may matter in receiving further education, since they are often among the entry requirements to higher levels of education, and can also work through the reinforcement of the students' self-consciousness with respect to their career options. Even if the skills measured by standardized tests do not matter regarding future earnings, the further levels of schooling accessed by higher test scores clearly do – at least through their signaling role. Besides, Hanushek (1986) argues that test scores appear to be valued in and of themselves, and educators, parents and decision makers also seem to value higher test scores. Although scores on the NABC are not used for selection into higher levels of education, they are explicitly designed to measure important skills.

There is no consent in the empirical literature whether class size reduction in the policy range of 15-40 students has any effect on achievement. Earlier papers had highly conflicting results both regarding the magnitude and the direction of the effect of various school resources, including the teacher/student ratio and class size (see Hanushek 1986 and 1997). On the other hand, most available studies with more convincing methodologies, such as using experimental data, natural experiments and instrumental variables, are more confident that the policy tool has a positive impact. The only large scale experiment called Project STAR was organized in Tennessee, and its evaluations generally came to the conclusion that class size reductions have strong positive effects on test scores (e.g. Finn and Achilles, 1990; Krueger, 1999). Quasi-experimental studies mostly came to similar conclusions (e.g. Angrist and Lavy, 1999); a notable exception being the paper by Hoxby (2000).

Another segment of the literature estimates the effect of various school resources on labor market outcomes. Card and Krueger (1996) provide an overview of the studies on how school quality affects future earnings, and also estimate the effects utilizing a "natural experiment" in North and South Carolina. The latter analysis relies on the historical difference regarding school resources (pupil/teacher ratio) between black and white schools and between the two states. They found that in South Carolina, where more school resources were diverted to white schools from black schools, black men in the 1900-1909 had 6 percent lower wages on average, while white men earned 8 percent more than in North Carolina. As the gap between the pupil/teacher ratios of the two states narrowed in subsequent cohorts, so did the differences in educational attainment and wages. In an earlier paper (Card and Krueger, 1992), the authors came to a similar conclusion: the effect of a year of education on earnings in a fixed labor market was higher for students who were educated in states with fewer pupils per teacher, higher average teacher pay, or a longer school year. By using the data of subsequent cohorts from the 1980 census, they could also control for state fixed effects, which results in even higher estimates of the impact of school quality. Heckman, Layne-Farrar and Todd (1995) performed a similar analysis on the 1970 and 1990 census data, and estimated larger effects for these years. They also extended the analysis by including regional aggregate demand and supply variables and allowing for differential school resource effects by level of education. While the inclusion of the former did not alter the general results, they found that that school resources had an effect on earnings only for those groups who had education level above high school. Krueger and Whitmore (2001) analyzed the longer term effects of the Tennessee STAR program on standardized test scores up to the 8^{th} grade and participating in a college entrance exam and found significant positive effects. Chetty et al. (2010) estimated the impact of STAR on the participants at the age of 27, and found positive effects with regard to multiple outcome variables, such as chances of getting into college, savings and home ownership. These analyses point to the importance of educational resources in future labor market outcomes. If such an effect indeed exists and it is partly a result of the difference in skills obtained during compulsory education, it may be possible to identify its sources by analyzing the data on student achievement.

In Hungary, the 1993 legislation on public education, which was still in force in the period examined in this thesis, prescribed maximum and average class sizes by categories of institutions and grades. The new law on education lowered the upper limit on class size in general, presumably partly as a result of the steadily decreasing number of students, although the relative drops in the number of students and teachers were similar in elementary schools (see Table 10 in Appendix B). A paper convincingly estimating the relationship between class size and achievement may inform both policy makers and the public about the effects of such policy changes.

To my knowledge, no previous papers were published that used the Hungarian NABC to study the much debated topic of class size. This database can be used both for cross-section methods on a large population of students and to apply panel methods to remove time constant unobserved fixed effects. After estimating a standard first difference specification, I use two other approach for the identification of the relationship. The first one is based on the method by Angrist and Lavy (1999), and uses the maximum class size rule in Hungary as an instrument for actual class size, which can be considered valid if certain conditions are met. I argue that the instrument is most likely not valid in the context described in the thesis. The other approach combines a first difference model with the fact that many students had to leave their schools due to the closure of these institutions, while many other schools are obliged to admit every student living in their district. Unfortunately, none of these methods rule out endogeneity convincingly enough. The only robust and statistically significant estimates are provided by the first differenced model with controls, but these are very small in comparison with the empirical literature.

The thesis is structured as follows. Chapter 1 first summarizes why the estimation of class size effects are problematic from a theoretical point of view, and then it provides an overview of the empirical literature dealing with the topic. Chapter 2 introduces the main characteristics of the Hungarian public education system. Chapter 3 describes the dataset and variables used, and present an exploratory analysis. Chapter 4 introduces the main empirical models and methods – first differenced OLS and two stage least squares – applied during the analysis, and demonstrates the potential drawbacks of these approaches. Chapter 5 presents the main results of the estimations. The last chapter concludes and discusses the potential for improvements.

Chapter 1: Literature review

1.1 Theoretical considerations

Theoretical modeling has not been a primary concern for economists trying to estimate the effects of class size on achievement; such works mainly reflected on policy concerns. The simple intuition that the more children are in a class the less time a teacher can allocate to each student could be estimated simply if the behavior of various stakeholders such as teachers, school management and principals would not respond to the relationship between the variables in question. By modeling the optimization behavior, it becomes obvious that serious endogeneity issues arise. I describe here a simple yet insightful model developed by Lazear (1999) focusing on how class size may enter the objective function of schools. Although his model assumes that schools are private and optimize profits, the main results of the analysis still apply to some extent if one assumes that society (receiving the incomes and bearing the costs) tries to maximize this profit. It must also be noted that the model cannot be used directly for estimation, and it has significant limitations due to its numerous simplifying assumptions; but it provides a few key insights into the problems of measuring the effects of class size. The model illustrates this effect as a problem of congestion, where students impose negative externalities on others by disrupting the instruction in the class. The probability that a student is not impeding his own or others' learning in a given moment is p. Then the probability that at any given moment no disruption occurs is p^n, where n is the number of students in a class; and $1-p^n$ is the probability that any student impedes the instruction. One student's such behavior disrupts instruction entirely for that moment, so in this sense Lazear's assumption of independence is not too restrictive. It must be noted that p should be very close to 1 when class sizes are realistic, as even $p = 0.98$ in a class of 25 students interrupts teaching 40% of the time. When no one impedes teaching, the probability that learning occurs by a student is q, so the amount of learning by the entire class is then $n\,q\,p^n$. If the value of one unit of learning is v, then a student's expected benefit from attending a class of size n is $v\,q\,p^n$. This can be conceived as the monetary value of a unit of learning. The model assumes homogeneity with respect to *p, q* and *v*, which is also one of its main limitations. If *Z* is the number of students in a grade, *m* is the number of classes and *W* is the cost consisting of teacher salary and the capital price associated with the classroom, an optimizing school would maximize the following profit function:

$$\pi = Z\, v\, q\, p^{Z/m} - W\, m \qquad (1)$$

where classes are of equal size $n = Z/m$, and *Z* and *m* are such that *n* is an integer, again by assumption. The first order condition with respect to the number of classes is then

$$-v\, q \frac{Z^2}{m^2} p^{\frac{Z}{m}} \ln(p) - W = 0. \qquad (2)$$

In order to have an interior solution, the second order condition is that

$$\frac{\partial^2 \pi}{\partial m^2} = v\, q\, Z^2\, p^{\frac{Z}{m}} \ln(p)\ \frac{2\, m+Z \ln(p)}{m^4} \qquad (3)$$

has to be negative, which is met if $2\, m + Z\, ln(p) > 0$. The implicit differentiation of (2) with respect to *p* yields:

$$\frac{\partial m}{\partial p}|_{FOC} = \frac{v\, q\, Z^2\, p^{\frac{Z-m}{m}} (m+Z\ \ln(p))/m^3}{\frac{\partial^2 \pi}{\partial m^2}} \qquad (4)$$

This expression is negative if the second order condition is met and *p* is close to 1 – that is, the numerator is positive. This shows that optimizing schools will *ceteris paribus* choose to have fewer classes in a grade if the propensity of students to disrupt is lower. On the other hand,

$$\frac{\partial m}{\partial q}|_{FOC} = \frac{v\frac{Z^2}{m^2}\ p^{\frac{Z}{m}} \ln(p)}{\frac{\partial^2 \pi}{\partial m^2}} > 0, \qquad (5)$$

so holding everything else constant, an increase in the propensity to learn results in a decrease in class size: since the gain by having less students in a class is higher, it is worth hiring more teachers. In this case, a decrease in observed class size further enhances the outcomes of more talented students, and the observed negative correlation is biased in the negative direction due to unobserved factors such as ability. Departing somewhat from Lazear's model and assuming that *p* is a function of *q* such that $p'(q) \geq 0$ will have the somewhat different implications (see Appendix A), i.e. due to the different levels of *q* or *p* in various classes and the optimization by schools, the estimated relationship is going to have a negative bias. The model also provides an explanation for the difficulty of finding any effect at all, for example "[i]ncreasing class size from 25 to 27 would reduce educational output per student by only about two percent" (Lazear, 1999: 12). It should be emphasized that many of the rather restrictive assumptions of the model can be very important regarding its exact implications, and one cannot infer the direction of the bias simply by relying on it.

In the real world, however, the managements of public schools are not the sole optimizing agents, and their behavior depends not only on the narrowly defined student characteristics and the costs associated with education. First, schools do not necessarily optimize aggregate outcomes and may consider equity concerns important. Second, the heterogeneity of school quality, student and household characteristics are likely to give rise to an equilibrium in which different students are sorting to schools of varying quality due to the interplay of these factors.[1] Thus, if one observes only the simple correlations between test points and school resources such as teacher student ratio or class size, one may observe either negative or positive relationship due to the omitted variables bias. For example, students with lower ability may be deliberately placed in classes with higher per capita resources to make up for their disadvantages. On the other hand, family background variables such as income and the concern of parents for their child's education also tend to correlate. Moreover, enrollment is often based on observed achievement even in some elementary schools, especially in the 6 and 8 grade high schools, as I mention later. Thus, children with better family backgrounds tend to get into schools with more resources both because the family is more

[1] A more complex model incorporating these elements was developed by Urquiola and Verhoogen (2009) for the liberalized Chilean educational system, see later in the thesis.

able to afford it (e.g. the moving or travelling costs) and because of the higher achievement of the child due to increased parental support. This bias is further exacerbated by the response of parents to the innate ability of a child: they may also tend to invest more in children who they think will gain higher return from education, or try to compensate for lower abilities by devoting more resources to less gifted children. Since innate ability is by definition an unobserved variable, it is not possible to control for the latter factor directly.

Another significant hardship in identifying the effects of the quantity of certain educational inputs on achievement is that the quality of education cannot be measured directly. One can use proxies such as the availability of tools, teacher experience, education and salaries which are reasonably considered to be highly correlated with quality, but these do not necessarily grasp the effectiveness of instruction. The quality of education is most likely correlated with other explanatory variables, like family background and class size. For instance, better educated or wealthier parents may have more information about school quality due to a higher level of social capital (e.g. connections), or simply more resources to enroll their children in better schools. Moreover, the quality of instruction is most probably correlated with class size and family background too: the demand for "good classes" is higher, but in the public school system both the class sizes and the number of classes are exogenously determined by the availability of public resources. If the number of good instructors and that of classes with high quality instruction is limited in a geographical area, and quality varies substantially across classes, the better ones are expected to enroll more students. In addition, the ratio of students with better family background will probably be higher in the better classes if more educated or wealthier parents have more leverage as to which class their children are being enrolled.

1.2 Empirical results in the literature

The so-called "Coleman Report" (Coleman et al., 1966) is considered to be one of the earliest analyses of educational production functions (Hanushek, 1986). The aim of the report was to study the distribution of educational resources in the United States by ethnic background. The report found that it was family background and the characteristics of other students that had the greatest impact on student performance, and differences in schools had little to do with achievement. The extensive criticism in response to the study partly focused on the production function approach, partly on the actual results. Hanushek (1986) provided an overview of the literature that deals with the relationship between educational resources and student achievement. Of the 112 separate estimates of the effect of student/teacher ratio, only 9 found statistically significant positive effects, while 14 of them displayed negative effects that were statistically different from zero. An update of the review found that 15% of the 277 estimates measured statistically significant positive effects and only 13% showed negative impacts, but the author's conclusion was again that the literature is inconclusive about the effect of school resources on achievement and teacher/pupil ratio in particular (Hanushek, 1997). The reviewed empirical studies often could not handle the specification problems due to the availability of only contemporaneous variables. Others estimated so-called "value-added" specifications, in which the difference in the dependent variable between two time-periods was on the left hand side, or achievement in period 1 is also added as an explanatory variable to the right hand side next to the *level* values of other explanatory variables. Applying this formulation, however, the "growth" or lasting effects of omitted variables can still bias the estimates (Hanushek, 1986).

The experimental project in Tennessee (Student Teacher Achievement Ratio, STAR) was explicitly aimed at measuring the effects of class size on student achievement and later life success. Altogether 11,600 kindergarten students and 1330 teachers were randomly assigned to classes of different sizes beginning in the 1985-1986 school year (Krueger, 1999). 13-17 students were enrolled in the small and 22-25 in the regular sized classes. In addition, there was an additional type of classes with regular size to which a teacher aide was also assigned. The students were asked to stay in the same class type for four years to which they were randomly assigned within the schools. The students sat for a battery of standardized tests at the end of every school year. The analysis performed by the internal team of researchers concluded that students in the small classes performed better than the ones in regular classes, while the performance of those with a teacher aide showed no difference from the latter. The above number also includes the new students who entered the schools during the years, and data have been collected on students through the 9^{th} grade. The results of Project STAR were estimated to be positive. Finn and Achilles (1990) found that first grade children who were in small classes had higher scores of 0.17 to 0.27 standard deviation on various tests compared to those who got into regular-sized classes without an aide (Mosteller, 1995). The results of a non-experimental pilot project in Wisconsin called Student Achievement Guarantee in Education (SAGE) were consistent with the STAR results according to its evaluators (Molnar et al, 1999). Schools participating in SAGE decreased their pupil-student ratios to levels between 12:1 and 15:1, which are lower than in the Tennessee experiment. This was accompanied by three other measures including the development of the curriculum, staff accountability and establishing "lighted school-houses" open all day long. Out of these four measures only class size reduction was implemented uniformly and immediately across the participating schools. Thus, the evaluation could focus on the effects of this intervention separately, and used comparison schools with student teacher ratios between 21:1 and 25:1 which were similar regarding other characteristics. Using OLS and a hierarchical linear model with the student and class level variables, they estimated that students gained significantly higher scores in all tested areas due to participation: the effect size of the class size reduction was about 0.2σ (Molnar et al 1999). It must be emphasized that these findings do not refer to per-student effects but the impact of the interventions.

Krueger (1999) analyzed the data from the STAR experiment using the ordinary least squares and the instrumental variables approach as well. Due to the occurrence of switching between classes and changes in class size during the years between kindergarten and 3^{rd} grade – the time span he analyzed – the author used initial assignment to a class type as an instrument for actual class type. According to the 2SLS results, a reduction of ten students is associated with a seven-tonine point increase in the average percentile ranking of students regarding SAT scores, depending on the grade. The estimates of pooled models including various class-level characteristics, like the fraction of students receiving free lunch and that of classmates who were in the same class in the previous year hardly change the beneficial effect of entering a small class. This effect is mostly attributable to a one-time jump though, as the coefficient estimated for the cumulative years spent in such a class are much smaller, and not always statistically significant. The value-added specification in which the author used achievement gain as the dependent variable reached similar conclusions, and thus results in smaller estimated effects due to "differencing out" the large initial gain in achievement. Comparing the effects by various sub-groups of students, he found heterogeneous treatment effects; e.g. disadvantaged and black students were shown to be affected more by attending a small class. In order to examine if the gain was attributable to Hawthorne or John Henry effects[2], he estimated the effect using the subsample of regular classes – where class size variation was likely to be a result of idiosyncratic factors – and obtained similar results.

Rivkin et al (2005) employed a semiparametric method to disentangle the effect of school teachers on the yearly achievement gains of students, and also estimated an education production function to reveal the impact of various observable student and school characteristics on achievement gains. They used a unique matched panel dataset from Texas, which includes all students and teachers in three cohorts in the mid-1990s. First, they estimated the part of test score variation within a school and a grade explained by the within-school variance of teacher quality, exploiting teacher turnover between cohorts of students. Accounting for student and school-by-grade fixed effects, their lower bound estimates imply that a one standard deviation increase in average teacher quality for a grade raises average student achievement in the grade by at least 0.11 standard deviations of the total test score distribution in mathematics and 0.095 standard deviations

[2] These occur when teacher respond to their initial assignment: the former applies to the case when small class teachers, the latter to the case when regular class teachers increase effort as a response to being assigned to one type of class.

in reading. Their estimates of the education production function include the achievement gain as the dependent variable. By controlling for student and school fixed effects, the authors addressed endogeneity concerns, for example compensatory resource allocations based on student performance and time invariant school characteristics that might be related to the included teacher and school characteristics. They also used school-by-year fixed effects in some specifications to account for systematic year-to-year changes in school factors. All of their specifications showed statistically significant relationship between class size and achievement, but the estimates including fixed effects were larger in absolute value than those without them. The impact was found to be stronger for mathematics than for reading, and decreased over time, but also remained different from zero in the 4th and 5th grade, which is only partly in accordance with Krueger's (1999) results. Moreover, the results showed important gains in teaching quality in the first year of experience and smaller gains over the next few career years, but little evidence of improvements after the first three years.

The seminal paper by Angrist and Lavy (1999) relied on class level data of the 4th and 5th grades in 1991 and the 3rd grade in 1992 in Israeli schools to analyze class size effects. The dependent variables were the average test scores of classes in reading comprehension and mathematics, while the explanatory variables included the percent of disadvantaged students in the school, total enrollment in the given grade in addition to the size of the class. Their simple OLS estimates had shown strong positive correlation between class size and achievement, but after adding control variables, these either lost their statistical significance or even turned negative. In the following regression discontinuity approach – which this thesis also utilizes – they used a discontinuous expected class size function (see below) as an instrument for class size. The authors found strong negative relationship between the class size and scores of fifth graders in all of the specifications using this method, and also for the reading comprehension test scores for fourth graders. On the other hand, the reported coefficients of class size for the sample of third graders were negative in all specifications, none of them were statistically different from 0. The estimated effect size of a one student decrease was 0.036-0.071 standard deviations for fifth graders and 0.017-0.019 for fourth graders in the distribution of class means.[3]

Many other studies exploited maximum class size rules in a similar way to Angrist and Lavy (1999). Fredriksson et al. (2012) found that smaller classes in the last three years of primary school (age 10 to 13) in Sweden were beneficial both in the short and the long term; students in larger classes scored significantly worse on cognitive and non-cognitive test at the age of 13; and their average score on high school tests (mathematics, Swedish, English) was also lower at the age of 16. The authors also found that those attending smaller classes had higher levels of completed education and wages at a later age (27 to 42). Urquiola (2006) used two identification strategies to estimate class size effects in Bolivia. First, he used data only on rural areas where there is only one grade per year, so class size is largely determined by the size of the cohort and the correlation with socio-economic status is less of a concern. Second, he applied the aforementioned IV approach. The results are statistically significant and negative in all of these specifications, but the IV estimates are substantially larger in magnitude. The method of using the class size rule to create an instrument was also applied for Danish data by Browning and Heinesen (2003), who found that class size reduction and reducing the number of pupils per teacher hour have large positive effects on educational attainment. It is noteworthy that the point estimates are significantly negative only after restricting the sample to the +/-3 intervals of enrollment around class-size discontinuity points (multiples of 24).

An instrumental variables framework was elaborated by Hoxby (2000) in order to estimate the effect of class composition and size on achievement using Connecticut school data. She utilized the natural variation in the population to create instruments: in small districts, school management cannot respond flexibly to changes in the size and characteristics of cohorts, thus the random

[3] It must be noted that this is calculated through dividing by between class standard deviation, which is necessarily smaller than the total standard deviation among pupils. This implies that using the total variance for calculation would result in somewhat lower calculated effect sizes (Angrist and Lavy, 1999).

variation translates almost directly into differences in class size. In another identification method, Hoxby (2000) modified the approach of Angrist and Lavy (1999) by first estimating district level "rules" regarding the number of classes in a grade as a function of enrollment in a cohort, and then using these to create expected class sizes serving as instruments. The level of analysis was the grade in a school, and the second identification method (maximum class size) is used in two ways: first treating data as though they were cross-sectional, and then making comparisons between adjacent cohorts in a first differenced model[4]. The overall conclusion of the estimations was in contrast with the findings from most studies with similarly credible approaches: Hoxby found that a reduction of class size in the range of 10 to 30 students does not have an effect on student achievement, as none of the more credible specifications resulted in statistically significant parameter estimates despite the small standard errors.

A statewide class size reduction program was launched in California in 1996 (CSR hereafter). The program reduced class sizes by roughly 10 students per class in kindergartens and schools up to the 3rd grade by setting the upper bound of class size to 20 students (it was 30 before 1996). The intervention had a large impact on teachers' labor market as well by creating almost 30,000 new teaching positions. The official evaluation found only limited evidence regarding the impact of program on student achievement (Bohrnstedt and Stecher, 2002). The student level analysis that controlled for student characteristics and compared schools that already participated to those that did not estimated small effects. On the other hand, a later school level analysis based on different exposures of cohorts to CSR and observed student characteristics did not find any impact. Another conclusion of the evaluation was that the program led to a decline in teacher quality. A paper by Jepsen and Rivkin (2009) focusing on the CSR used differences in class size and specific teacher characteristics by school, grade, and year to identify variable effects on achievement in a fixed effects framework. According to their estimates, reducing the number of

[4] The latter is called within-school regression discontinuity method, as it focuses on enrollment changes in a school between cohorts that trigger a change in the number of classes.

students per class by 10 results in an increase of mathematics and reading achievement by roughly 0.10 and 0.06 standard deviations within the school average test score distribution. They also found that the increase in the shares of new and not fully certified teachers partially offset the beneficial effects of class size reduction.

Chapter 2: The Hungarian educational system

The basic level of Hungarian education spans from the 1^{st} to the 8^{th} grade. Most students attend primary schools until 8^{th} grade, although there is a significant minority whose parents opt to take their children to special high schools after 4^{th} or 6^{th} grade. It is widely known and supported by the NABC data that the student composition of these special classes – the so-called 8 grade and 6 grade high schools – is significantly above the national average both in terms of the socio-economic background and achievements. The differences between various types of schools can be inspected by having a look at average achievement (see Tables 11-13 in Appendix B). As expected, average scores on both tests increased by 8^{th} grade. It is also apparent that there are significant differences between the average test scores of various types of institutions. This again tells us about the characteristics of Hungarian public education system, where the "traditionally" good schools select the high achieving students. Family background characteristics measured by the "family background index" are also quite different across school types.

Elementary education in Hungary is characterized by a decreasing number of students and teachers, which is accompanied by a declining number of school sites and classrooms as well (see Table 10 in Appendix B). The number of students decreased by 15.6 percent between 2001 and 2012, but this did not lead to smaller student/class and student/teacher ratios, which hardly changed in the period: the former revolved around 20, while the latter was in the close neighborhood of 10 throughout the period. The number of students in the 5^{th} to 8^{th} grades, which includes the sample examined in this thesis, showed a similar pattern although the decrease in their number was even stronger, reaching 23.34 percent.

Many studies point out the high selectivity of the Hungarian educational system; for example Csapó et al. (2009) show that the proportion of the variance of various standardized test scores explained by differences between schools is very high in international comparison. Based on the IEA data[5] the selection of students starts already at the beginning of primary school, and by the 8th grade 30% of the variance of mathematics scores can be explained by differences between schools. This ratio was essentially the same according to the official analyses of the 2011 and 2013 NABC data (the subjects of this thesis): about 30 percent of the variance of mathematics and reading comprehension scores was explained between-school variance in the 6th and 8th grades (Auxné et al, 2012 and Balázsi et al, 2014). Significant differences were observable between the average scores of settlement types; students in small municipalities (villages) had lower scores on average on both tests of about one half of a standard deviation than pupils in county centers.

Kertesi and Kézdi (2009) used the data of the 2006 NABC measurement to examine between and within school segregation and the difference in the level of segregation between municipalities. After calculating segregation indexes for the Roma and the disadvantaged, they found that its level is higher for the former group, and between school segregation is much stronger than segregation within schools and between classes. They also found that the availability of other schools in the geographical area (microregion or town) increases the possibility of attending a school in a different district. Their regression analysis showed that between school segregation is significantly affected by the number of schools and the ratio of Roma students in the area, but controlling for other factors, the ratio of disadvantaged students did not have an effect on the segregation index of the area.

The educational law in force determines the maximum number of students that can be enrolled in a class. Until September 1, 2013 this was 30 students for the 5th to the 8th grades, which was allowed to be exceeded by 20 percent if only two classes were launched in the beginning of the school year.[6] It is stipulated in the text that the school needs to ask for the permit of the competent educational bureau, the parental and the student organization in order to implement

[5] This is the main data source they used, the abbreviation stands for the International Association for the Evaluation of Educational Achievement

[6] These are described in Annex III of Act LXXIX of 1993 on public eduxation.

such an increase. Average class sizes are also specified for various grades, but the text of the law also includes that this does not have to be taken into account. It is important to note that maximum class size varies by grade: for the 1st to 4th grade it was 26 (+20%), which makes it less likely that students are enrolled up to the maximum level in 6th to 8th grade.

I will argue in the next chapters that the segregated nature of Hungarian public education makes it much harder to estimate the effects of class size on achievement, but this applies of course to school resources in general. The reason is that differences between the achievements of students in schools across various dimensions such as type of education, geography or the physical condition of the school site (see e.g. Auxné et al, 2012 and Balázsi et al, 2014) most likely reflect the socio-economic backgrounds of the students as well.

Chapter 3: Data and variables

The estimations are based on the research database of the National Assessment of Basic Competences (NABC) provided by the Hungarian Bureau of Education (Oktatási Hivatal). The NABC tests are administered in the end of May every year focusing on basic competences in two subjects, mathematics and reading comprehension. Participation is compulsory for every basic and secondary educational institution that has 6th, 8th or 10th graders. The yearly datasets are available with unique student level identification numbers since 2008, which allows constructing a panel database for students in the grades mentioned. From that year on, the tests for various grades have been using a common scale, which means that the tests assess the same set of skills and a subset of the questions is the same in all the three grades. The objective of the latter is to make the results comparable both across grades and years. The test scores are standardized based on the 6th grade results, which are set to have an average score of 1500 and a standard deviation of 200 (Oktatási Hivatal, 2010). In addition to the test scores, various student, site and institution level variables are included in the dataset. With a few exceptions,[7] all students have to take part in the NABC.

I excluded students with special educational needs (SEN) and those struggling with integration, learning and behavioral problems (ILB) from the sample. The reason for this was that as these students are officially declared disadvantaged, specialized instruction methods and on significantly smaller class sizes may apply to them. The analysis by Boozer and Rouse (1995) points to similar concerns in the context US schools, mentioning that remedial and special education classes possibly produce test score gains with a different production technology. Although including dummies and interaction terms may reduce the bias resulting from these, I decided to exclude these students from the models for the sake of brevity and the better comparability of fixed effects specifications with other models.

[7] These include students with disabilities, autism or injuries that prevent them from writing the test.

Due to computational constraints the thesis analyses the test results of one cohort, the students who were in 6th grade in 2011 and 8th grade in 2013. I restrict the estimations on these two years, as most students go to high school after finishing 8th grade, and one of identification strategies focused on those who did not change school between 2011 and 2013. A significant attrition can be observed: in 2011 (6th grade) the cohort included 94,047 observations, which decreased to 89,913 by 2013. It must be noted that both years contain observations that are not present in the other: most probably these students either had to repeat grade 6 or 7, or migrated to/from another country. One disadvantage of the dataset is that although they contain class identifiers for every school in every year, these are not consistent across years. Thus, it was possible to create unique class ID's within a given year, but not across years. Fortunately, this did not pose a serious problem during the estimations.

The family background index (FBI) is calculated based on the answers to the student questionnaire that accompanied the competence measurement tests. It is important to note that these answers were given by the students and they are not authenticated by comparisons to administrative data. The main principles behind the choice of the calculation method of the family background index were explanatory power and data availability (see Auxné et al, 2011 and Balázsi and Zempléni, 2004). The effects of numerous background variables on achievement had been calculated in a linear model using all grades and both tests in the year 2006. In the end, only those variables were included in the index that had significant effect on achievements according to the model, and the content of the index did not change in the subsequent years. These variables are: number of books at home, education level of parents, computer at home, and the number of own books. Unfortunately, the variable is not available for a large number of students, and its availability is perhaps not independent of the dependent variables, as there is a some negative correlation between the dummy for missing FBI and scores (-0.0997 for reading comprehension and -0.0921 for mathematics in the pooled sample). The negative correlation of the dummy with class size is much smaller (-0.0364), but regression estimates including FBI as an explanatory variable are distorting the results. This problems is not addressed in this thesis, but I expect it to cause a downward bias.

The class size and enrollment variables were constructed by me. While the former variable was not available, the latter was provided by the school principals. Since even absent students are included in the database without score, the calculation of class sizes was accurate most probably, although there are a few classes with only one student; they are most likely private students. For the sake of consistency, I used the enrollment variable created by adding class sizes in the school, which was not always the same as the enrollment count provided by the principals.

For the last empirical approach I calculated the number of students who had to leave their school after 6^{th} grade due to the closure of their institution by municipalities and by postal codes. I identified these students by 1) observing whether an institution has disappeared from the database by 2013 and 2) by comparing this to the list of discontinued institution available on the website of the Bureau of Education. If an institution was absent in 2013 and it was declared as discontinued by the Bureau in the school year 2011/2012, I considered the 6^{th} grade students to be forced to leave their former schools. A problem with the approach could be that even if an institution was officially discontinued, a new one may have been created with exactly the same teachers and students at the same place (a legal successor), which I could not identify based on the available information. Measuring the effect of class size change only with respect to students that did not change school handles this problem, as those schools where all students are class changers are removed from the sample.

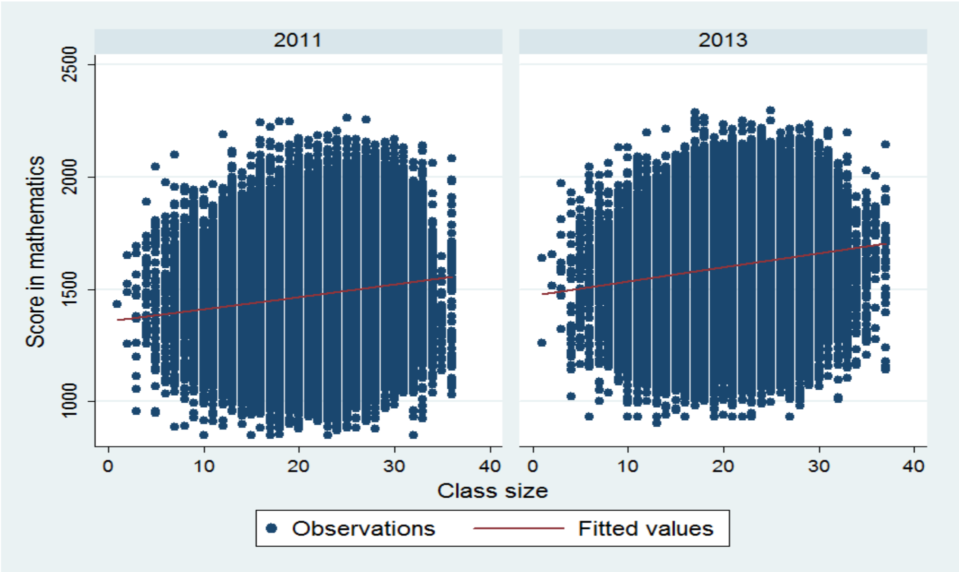
As it can be seen in Table 1, both class size and family background index are positively correlated with scores on the two tests. The high correlation with family background index is a result of the construction of the variable (for scatterplots, see Figures 5-6 in Appendix B). Although the elements of the index can be considered predetermined, due to omitted factors such as selection into better schools and the age[8] one should not infer causality based on this. On the other hand, the correlation with class size is also positive, which is against the intuition, and also draws attention to the inadequacy of simple pairwise correlations for causal inference. For example, it may suggest that schools place better behaving children in larger classes as predicted in the model described in Chapter 1; or it can be the result of higher demand for places in "elite schools", primarily by more assertive and wealthier parents, which generates larger classes in these schools (see the further discussion in the next chapter). Such a positive relationship between the two variables is also apparent when fitting a line on the scatterplot of the data (see Figure 1 here and Figure 5 in Appendix B), which is estimated to have a slight positive slope. Unsurprisingly, scores in mathematics and reading comprehension are highly correlated.

Table 1. Sample correlation coefficients of the main variables

	2011				2013			
	FBI	Class size	Math score	Reading score	FBI	Class size	Math score	Reading score
FBI	1				1			
Class size	0.235	1			0.207	1		
Math score	0.436	0.141	1		0.482	0.162	1	
Reading score	0.499	0.167	0.693	1	0.516	0.153	0.739	1

[8] See Hámori and Köllő (2011), who analyzed how the age at which a child enters first grade affects achievement in the Hungarian context.

Figure 1. Scatterplots of scores in mathematics and class size



Continuing the preliminary analysis, the seemingly positive relationship remains if one tries to evaluate the effect of class size and family background on scores in a simple OLS framework. The estimated coefficients are highly significant but small in magnitude; they suggest that increasing class size by five pupils would result in an increase in scores by about 0.03-0.05 standard errors. The positive estimates are in contrast with intuition, but they can be easily explained by the model and the possible other mechanisms described in Chapter 1. First, if schools optimize with respect to class size, larger classes will consist of students who are less likely to disturb instruction – a characteristic that is likely to be correlated with ability. Moreover, even in the absence of optimizing behavior which is more reasonable to assume in the case of private schools, the demand for places in better schools can be larger, which may lead to higher enrollment in these. Thus, due to endogeneity and omitted variables, these estimates are likely to be highly biased and inconsistent, which I am going to attempt to deal with in multiple ways, as presented the next chapter. Further sources of endogeneity emerge in the Hungarian context, which are also going to be discussed in the next chapter.

Table 2. OLS regressions of test scores on class size and family background index

	Mathematics, 2011	**Reading, 2011**	**Mathematics, 2013**	**Reading, 2013**
class size	1.192*** (.319)	1.485*** (.234)	2.139*** (.290)	1.437*** (.245)
family background index	84.474*** (1.319)	89.936*** (1.043)	92.945*** (1.265)	98.070*** (1.149)
R^2	.192	.251	0.235	.267
N	64046	64046	56163	56163

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01.*

Chapter 4: Empirical model and estimation methods

I model student competence scores as a function of observable and unobservable student, class, school, grade and year level characteristics. Since I use only one cohort for the estimations, grade fixed effects and interactions correspond to year effects, thus I only include the former in the model. The achievement for a student in a given grade is produced as follows:

$$A_{icsy} = \alpha + \beta' X_{icsy} + \gamma' C_{csy} + \delta' SG_{sy} + \epsilon Y + \varepsilon_{icsy} \qquad (6)$$

Where α is a constant; X_{icsg} denotes a vector of observable student level family background variables for student *i* in class *c*, school *s* and grade/year *y*; C_{csy} is a vector of class characteristics including class size, average family background variables and competence level of other students; and SG_{sy} denotes school and grade level variables, such as the location of the school and enrollment in the school for the given grade. *Y* denotes year and in this context it is equivalent to grade. Finally, ε_{icsy} is the term that captures the unobserved characteristics at all levels. If it was uncorrelated with the observable explanatory variables, the model could be estimated consistently by ordinary least squares. Based on the arguments in Chapter 1, it is reasonable to assume that the terms incorporated in ε_{icsy}, such as the student's skills and the quality of instruction, are determined jointly with observable variables. In order to model the error term, I assume the usual additive form:

$$\varepsilon_{icsy} = \theta_{ics} + \vartheta_{cs} + \pi_s + \varphi_{cy} + \tau_{sy} + \omega_{icsy} \qquad (7)$$

Here θ_{ics} is a time invariant fixed effect that can be thought of as innate ability; ϑ_{cs} is time invariant class-level fixed effects such as classroom, teachers or method of instruction that do not change in the time-span examined here; π_s is school specific unobserved effects constant in time; ϕ_{cy} denotes unobserved time variant class-specific shocks such as change in teacher compositions; τ_{sy} is a school specific shock such as a change in school management, and ω_{icsy} is the idiosyncratic error. Ordinary least squares estimation that uses a rich set of controls does not account for the presence of constant and time varying unobserved effects that are correlated with the explanatory variables, therefore estimating the population regression coefficients with OLS would result in biased and inconsistent estimators. On the other hand, the application of fixed effects or first differenced specifications is also accompanied by the assumption that the idiosyncratic error is uncorrelated with the observed variables included in the model, which is a condition for the consistency.

Unfortunately the first differenced or the fixed effect estimation strategy cannot deal with the issue of year (grade) specific shocks on the individual, class and school/grade level. Although it can be assumed that after differencing out time constant individual fixed effects the idiosyncratic errors ω_{icsy} are not correlated with the change of observables, this is not necessarily true. One can think of e.g. "rewards" provided by families to the child for good grades, which are correlated with family background (e.g. through the valuation of good grades and wealth) and provide incentives for increased efforts. This can especially appear in the grades examined here, where the approaching high school applications may induce parents to apply such incentives. School and class specific shocks that may be similarly correlated with increased efforts are for example school or class competitions, the effects of which may appear on all levels. Schools and classes with e.g. higher enrollment, wealthier student composition and better teachers may be more likely to launch competitions that may uniformly increase achievement, but more gifted students with better grades are usually supported by teachers to a larger extent. Including such individual specific variables may therefore result in a bias. On the other hand, the change in class size between period t and $t-1$ is possibly correlated with shocks such as the change in the quality of teachers in period t. The bias caused by these confounding variables depends on the relationship of class size with unobserved family background and ability.

In contrast to (6) and (7), it is not necessarily true that achievement in period t is affected by the past level of the explanatory variables to the same extent as it was in period $t-1$. This implies not only that omitting the past levels of these variables causes bias, but also that a fixed effects framework cannot deal with them either. This problem is very similar to the time changing unobserved effects problem if past values of the explanatory variables are not observed, but they are serially correlated. This implies that the "true model" would look like the following for any given year *y*:

$$A_{icsy} = \alpha + \sum_{t=0}^{y} \beta_t{}' X_{icst} + \sum_{t=0}^{y} \gamma_t{}' C_{cst} + \sum_{t=0}^{y} \delta_t{}' SG_{st} + \epsilon_y Y + \varepsilon_{icsy} \quad (8)$$

I assume that the relationship among the variables is stable in time, so coefficients are the same for any *y*, e.g. $\beta_{y,t,1} = \beta_{y-1,t,1} = \beta_{k,t,1}$. First differencing (8) and changing the indexes of the time periods and the corresponding coefficients, the equation can be written as:

$$A_{icst} - A_{icst-1} = \sum_t \sum_k \beta_k{}' (X_{icst} - X_{icst-1}) + \sum_t \sum_k \gamma_k{}' (C_{cst} - C_{cst-1}) + \sum_t \sum_k \delta_k{}' (SG_{st} - SG_{st-1}) + \epsilon_y - \epsilon_{y-1} + \beta_0{}' X_{ics0} + \gamma_0{}' C_{cs0} + \delta_0{}' SG_{s0} + \omega_{icsy} - \omega_{icsy-1} \quad (9)$$

If only two periods are observed (as in this case), and the between-period changes are autocorrelated, a first differenced regression yields biased and inconsistent estimates. I find it plausible that past changes only have limited effect on present changes, but they are significantly correlated with an important source of bias: the initial values of the variables (these can be thought of as equivalent to first grade values). Another major bias stems from idiosyncratic errors correlated with the explanatory variables (see the examples above). The correlation of changes with the initial values can be especially problematic in the case of class size, where equalization may occur throughout the years. For example, if due to sorting, pupils with better background were enrolled in smaller classes, newcomers in later years may be placed into these classes by the school management. This results in a positive correlation between past values of family background and class size change and an upward bias.

Based on intuition and the models and empirical findings described in Chapter 1, I expect that class size has a negative effect on achievement, and a reliable identification strategy should find small and negative per-student effects. I use three empirical strategies to estimate the effect of class size on achievement in the model(s) outlined in (6) to (9). The simplest approach would be to consider only towns so small that they have only one school and one class in each grade. Although the dataset contains variables about the size of the town where the school is situated and the presence of similar schools in the neighborhood, there is a significant problem with this method. Kertesi and Kézdi (2005) argue that middle class parents are more likely to take their children to better schools in other municipalities for many reasons. First, the costs of commuting are not affordable for poorer families, second, wealthier parents have on average better information regarding the quality of schools, and third, the return of attending a better quality elementary school is higher for wealthier children. However, by estimating a **first differenced model** using only the data of students who did not change school during the two years between the two measurements, one can eliminate the time constant fixed effects that are most likely to be correlated with the observed variables, including class size.[9] Besides class size, the explanatory variables of this model included the composite index of family background which was constructed to include any important background variables that were widely available, enrollment, the average family background index of other students in the class and the average test score of other students. Including other observed variables – the status of being multiply disadvantaged, and the ratios of students with specific education needs, multiple disadvantages and integration and learning problems – did not change the main results; therefore I omitted them from the main specifications. This specification yields unbiased results only under the assumption that (1) there are no time varying factors that are correlated with both the change in any of the explanatory variables and achievement, (2) the changes between time periods and the initial values are uncorrelated. It must be mentioned that there is nonrandom attrition: those students who appear in the database in only one year are supposedly the ones who had to repeat a class, and therefore are lower achievers. The first differenced models do not account for them, which is problematic if their achievement and non-inclusion in one of the years is partly a result of class size; the magnitude of the effects will be

[9] I assume that between class movements within schools were not significant.

underestimated. There are 4275 students who only appear in the year 2011 and 1478 who are present only in the 2013 data. The correlation between the dummy for appearing only in the year 2011 and class size is slightly negative (-0.0614), which is counterintuitive, but possibly the result of the school's behavior of placing such problematic pupils in smaller classes. The correlation of class size with the dummy for appearing only in 2013 has almost the same size (-0.0611).[10]

The problems that emerge in the case of estimating a first differenced model can be eliminated using appropriate instrumental variables that are uncorrelated with unobservables and potential outcomes, but still have an effect on class size. The **cross-sectional instrumental variables approach** of Angrist and Lavy (1999) uses the administrative rule – termed Maimonides' rule after the twelfth century rabbinic scholar – that specifies the maximum class size at 40. The "fuzzy" regression discontinuity (RD) design in this context relies on the fact that expected class size calculated using the enrollment in a given school and grade and the maximum class size rule jumps discontinuously when the cap is reached, thus it is reasonable to assume that just at the point of discontinuity it is uncorrelated with the confounding variables that are otherwise correlated with enrollment. The class size function looks as follows:

$$C_f \equiv \frac{enrollment}{int\left(\frac{enrollment-1}{class\ size^{max}}\right)+1} \quad (10)$$

The rule is not deterministic since there are non-compliers, hence the RD approach is termed fuzzy. Although the number of classes in a given grade is an endogenous decision made by school leadership, the majority of them are assumed to consider the rule binding. Therefore, an increase in the number of classes, and thus, a decrease in average class size around the maximum are exogenous to their optimization. The further assumption is that although enrollment may be correlated with unobserved variables that are in turn affect achievement, it is a continuous function

[10] Surprisingly, both are significant statistically at the 5% level, the small magnitude of them suggest that their effect of the estimation is not large.

of such student/family characteristics; therefore the latter supposedly do not differ significantly on the two sides of the class size cap. Angrist and Lavy (1999) based this method on the work of Campbell (1969), arguing that a smooth function of the running variable – enrollment in our case – can be still controlled for in a 2SLS regression that employs expected class size as an instrument for actual class size. The authors and Hoxby (2000) note, however, that with cross-sectional data the approach works only in the close neighborhood of discontinuity points due to parents' endogenous choice of schools based on the usual realizations of class size in given schools in the past. In the neighborhood of these points, however, the underlying population of students can be assumed to be very similar (if all schools apply the rule in the same way).

Another important assumption for consistent estimation is that the class size rule is not exploited purposefully, and its application is not different across various groups. This statement partly corresponds to the independence condition outlined in Angrist and Imbens (1995), which states (among others) that potential outcomes must be independent of the variable considered as an instrument. If this condition is not met, e.g. those students that gain more by complying with the rule are more likely to benefit from it, the instrument is not valid. This problem is especially relevant in the context of liberalized school markets with varying quality and resources among schools, such as the one in Chile. Urquiola and Verhoogen (2009) elaborate a theoretical model for the Chilean system with multiple types of schools, households differentiated with respect to income, and sorting of households among schools. Their model incorporates the idea of Lazear (1999) that the larger the class size, the less teacher attention is available for each individual student. Two equilibrium results of the model are empirically testable: first, there is an approximately inverted-U relationship between class size and average household income; second, schools may stack at enrollments that are multiples of the class size cap, implying discontinuous changes in average household income with respect to enrollment at those points. The stacking generates discontinuities in the relationship between student characteristics and enrollment around the cutoff points, which violates the smoothness assumptions underlying the RD approach. The prediction of the model is that students on the right of the cutoff will come from wealthier households, which is also supported by their empirical investigation of schools subject to class size caps. Urquiola and Verhoogen (2009) show both by regression analysis and visually that parents' schooling and income display a statistically significant increase at the first enrollment cutoff. Moreover, the inclusion of family background variables in the IV specifications causes the class size variable to decrease in magnitude and lose its statistical significance, which is an additional indication that the smoothness condition is violated.

The conditions specified for the RD approach are not entirely met by the Hungarian educational system. It is reasonable to assume that municipalities have incentives to enroll as many students in a class as possible, as they receive a normative subsidy by the number of students enrolled in their schools (Kertesi and Kézdi, 2005).[11]The reason is that a large part of the school income is spent on the fixed costs (e.g. infrastructure, teacher salaries), thus they can decrease the average costs per capita by increasing the student number up to the limit. Such incentives results in a "competition of quantity" between schools. On the other hand, there is also a "competition of quality" for higher achieving students, as this enables teachers to secure better working conditions and achieve success and respect in their middle class social environment. The initial conditions regarding the quality of education[12] in a school is of central importance, since historically better schools will attract better students, which further enhances the quality of school by being able to hire better teachers for the reason that the value of teacher salary is higher in them as teachers have to exert less effort to reach a certain level (Kertesi and Kézdi, 2005).

Another source of inconsistency can be the varying behavior of schools by type, which can be especially relevant regarding primary (and secondary) education in Hungary. An important

[11] This has changed to some extent due to the change of maintainer rights in early 2013, but this is not likely to have resulted in within year changes.
[12] These initial conditions in Hungary were mainly determined by the geographical area, as children had to enroll in their school district until the end of socialism.

feature of the system is that there are two types of public schools regarding the area of enrollment. District schools are obliged to enroll every child applying from their area, and may also enroll students coming from other district if there is remaining capacity. On the other hand, there are primary schools with national level enrollment. These do not have to enroll every student applying from the district in which they are located, but the number of classes to be launched is pre-determined along with the number of students to be admitted. This maximum number may be different – on a case by case basis – from the national rule for first graders, which is also lower than the maximum class size of 6-8 graders. Therefore these schools may choose the best students from the pool of applicants, yet have large average expected class sizes. Unfortunately, the NABC dataset does not contain information about these characteristics of the schools, so sorting between types may result in biased estimates if one uses predicted class size as an instrument. The problem is not exclusively of schools with national and district level enrollment though. As noted above, the highly segregated educational system results in the commuting of wealthier – and thus on average higher achiever – students between municipalities, where it is up to the local school whether they are admitted. This makes for these historically higher quality schools easier to increase class size up to a limit in order to reduce per student costs. In addition, they can also enroll higher achieving students on average, as they can select from among the commuting students based on their skills and family background – both of which are highly correlated with the dependent variable, yet not entirely observable in the data.

Beginning with the school year 2007/2008 the schools that cannot admit every applicant in the 1^{st} grade have to organize a lottery, which alleviates selection. Unfortunately, the cohort analyzed in this thesis enrolled in 1^{st} grade in the 2005/2006 school year. It must be noted, however, that initial randomization does not inhibit selection due to other factors such as place of residence, resources committed by parents to commuting and student reallocation in later years. For example, if the pool of applicants participating in the lottery has higher potential outcomes on average, the selection problem still persists.

A further possibility is that two effects are present simultaneously. In small but wealthier districts (with higher achieving students on average), there may not be a constraint on starting new classes, as there are enough resources to hire new teachers. In poorer areas, this may not be the case, and students may be admitted above the maximum class size if the rule is not strictly enforced. Meanwhile, better schools may respond differently to the rule, as they may start new classes well before reaching the cap. Considering larger municipalities and schools with national level enrollment, the converse may be true: it is the better schools that enroll students up to a maximum and do not launch new classes, while other schools may abide by the rule if they can. Thus, two sources of bias can be simultaneously present, leading to the invalidity of the instrument.

A peculiarity of the class size rule in Hungary noted above is that it varies along with the number of classes in a school. Since the approach applied here does not need the exogeneity of the number of classes, nor enrollment, only that of the application of the rule around the cutoff points, this does not hinder this line of analysis.[13]During the estimations I used both class size rules, i.e. implemented the 2SLS method with the uniform rule ("soft rule", or rule 1) and the "hard" rule (or rule 2) as well. In the former case, I assumed that class size is capped at 30 pupils; while under the "hard" rule the maximum number of students in a class is 36 up to an enrollment of 72, which corresponds to two classes with the extended size. It is required for identification that the drops in class size as a result of an increase in enrollment are due to the class size rule and not to other factors that are correlated with enrollment. I consider, however, using the "hard rule" potentially more reliable. The reason is that abiding by the soft rule may well depend on the potential outcomes, e.g. middle class parents would not enroll their children into classes that are known to have classes larger than 30, and also would not agree to a later increase above that level. Rule 2, in contrast, is more stringent in the sense that class size (theoretically) cannot be increased above that level. Thus, the application rule 1 itself can be considered a choice variable jointly determined by

[13] It is reasonable to assume that the decision on which class size rule to apply is simultaneous with the decision regarding the number of classes.

the school and parents depending on the level of enrollment, while the application of rule 2 is not determined by the actors – although its validity can still be affected by sorting. Thus, expected class size based on rule 1 does not qualify as an instrument if the rule can be affected by the agents.[14] The results of using this rule are in the Appendix for reference, but the main text focuses on rule 2.

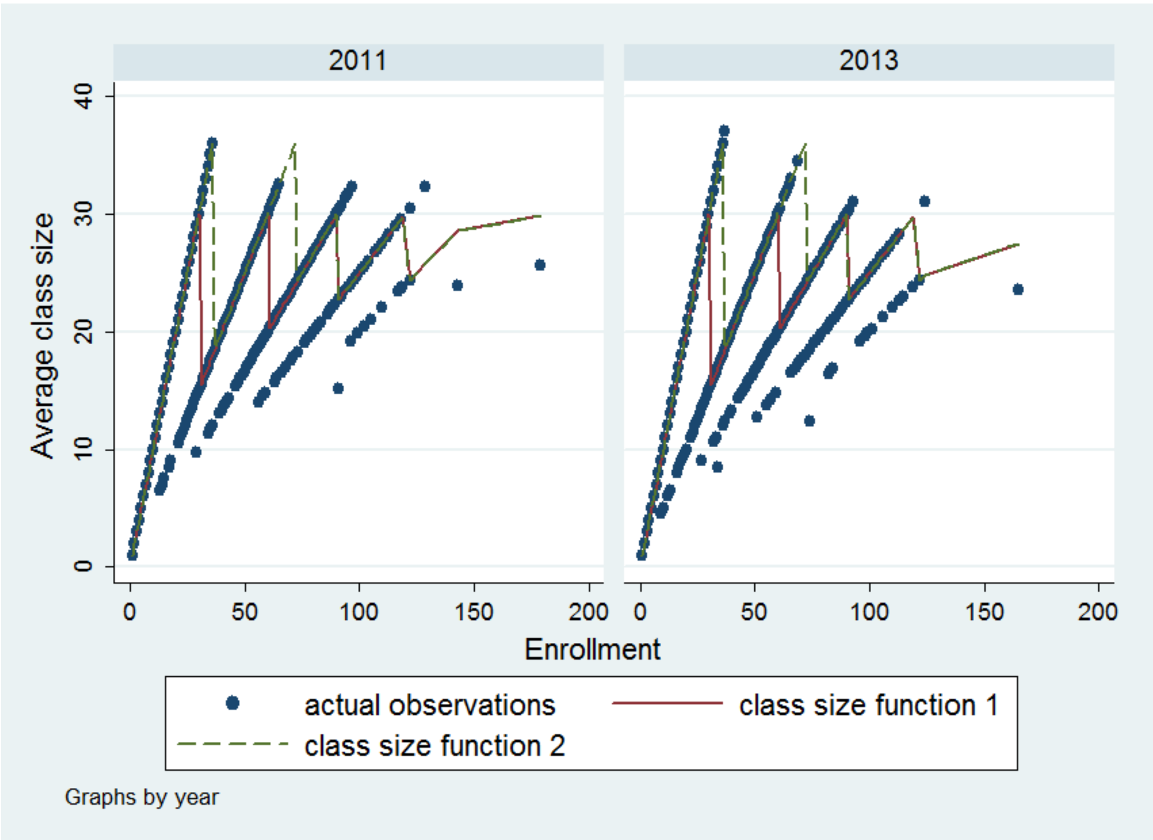
It must be reemphasized that the class size rule was different in 1st grade than in the 6th and 8th grade I examine here, so after the change of rule (5th grade) new enrollments may also have been endogenous for the above reasons. Parents had plenty of time to search for another school if the initial enrollment was resulted in large classes. Even if they did not change school during the years, wealthier parents may have been more effective in preventing an increase in class size, so newcomers were more likely to be placed in classes with less wealthy background. The contrary may also be true, i.e. better schools may have faced a larger demand, which in turn may have resulted in larger increases in class size.

As it can be observed in Figure 2, schools (at least in terms of the average class size) seem to mostly follow the rule of maximum until the enrollment of 90 students, but above that, some schools exceed it. Moreover, many schools start a new class well before reaching the class size rule, a fact that is against the prediction that schools are incentivized to enroll new students until reaching the cap. For example, there are no schools with an enrollment of 72 that reach the average class size of 36. The most likely reason for this is "historical": the number of classes in 6th grade is the same as 4th to 5th grades, where class size caps are lower. Most children do not change school, so there is no need to start new classes for newcomers due to the increase of the maximum in 5th grade. Moreover, the number of students was decreasing in the period, so schools may have tended to launch the same number of classes they used to in previous years until the number of students

[14] Neither the independence, nor the monotonicity conditions in Angrist and Imbens (1994) are met in this case for causal inference.

reaches some minimum value. A further reason is institutional/personal even though there may be random changes in the number of students, the school management cannot or do not want to dismiss teachers. These observations can be problematic in an instrumental variables framework due to the possible low relevance of the instrument (class size rule) in explaining actual class size. In the terminology used by Angrist and Imbens (1995), 2SLS estimation identifies the local average treatment effect (LATE), the effect on those whose treatment status is affected by the instrument. There are possibly quite few schools in the neighborhood of discontinuity points to which this applies.

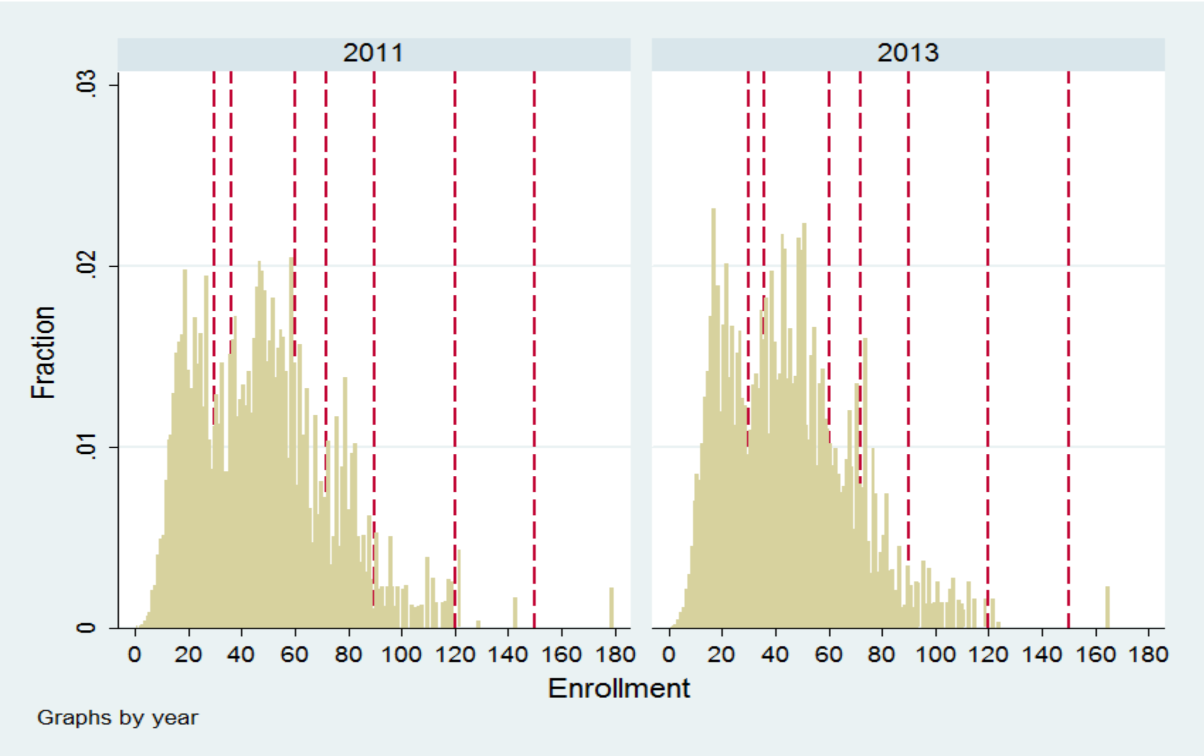
Figure 2. Predicted and actual average class size at school sites.



Note: the difference between the 2011 and 2013 lines for predicted class size is due to the calculations being based on the sample, and there were only two schools with enrollment in the neighborhood of 150.

In order to check whether sorting takes places around the enrollment caps I apply a simple analysis similar to that of Urquiola and Verhoogen (2009) and Fredriksson et al (2012). When inspecting the fraction of students in schools with various enrollment levels in Figure 3, one should look for sudden drops in the fractions of enrollment around the values associated with the caps. Although the figure is based on student level data, there still should not be discontinuous drops around these points if enrollment in various sizes of schools is a smooth function of continuous exogenous variables. It seems that there is a drop in their number to the left of the level of 30 and 36 in 2011, which may suggest that a significant portion of schools actually try to fill their classes, while the enrollment in other types of schools is in accordance with the rule for the 1st to 4th grade (see Chapter 2). The latter explanation is based on the plausible assumption that the number of schools in an area is determined by the time persistent size of cohorts. The jump before the enrollment of 36 may be contributed to a similar division of schools by type; although the class size rule for previous grades does play a role in the explanation (this enrollment size is associated with two classes in lower grades). On the other hand, the former (and smaller) jump is not present in year 2013, which may be associated with the response of students (parents) to large class sizes. All in all, it seems there is some kind of stacking at the enrollment level of 36, but the factors behind it remain hidden.

Figure 3. Fraction of students in classes with given enrollment sizes.



Note: enrollment levels of 30, 36, 60, 72, 90, 120 and 150 are shown by the dashed lines, which are the cutoffs for the class size rules

Table 3 contains the results of regressing the family background index on dummy variables indicating whether a student was attending a school with an enrollment level in the specified segment. When examining only the +/- 5 neighborhoods of the points of discontinuity, the positive coefficients show that being above the first two cutoffs in year 2011 is associated with a family background index 0.140 – 0.175 standard deviations higher. Examining the full sample and including dummies for the neighborhoods shows that being above the first cutoff in year 1 is associated with a 1.9 standard deviation higher FBI than being just below it in 2011. The results also show that being in the +/-5 neighborhood the first and thirds cutoff is correlated negatively with the FBI. The 2013 data, however, do not show such an unequivocal pattern, as being above the cutoffs is not significantly associated with higher (lower) FBI. This shows that if the class size rule strictly holds, pupils with better background were more likely to attend smaller classes due to being enrolled in schools just above the threshold in 2011. Since background has a significant effect on achievement and is possibly correlated with unobservables that have similar effects, these findings suggest that selection took place with respect to potential outcomes. It is puzzling however, that such differences were not shown for (mostly) the same students two years later – two possible explanations are higher level attrition from bigger classes, and an endogenous "equalizing mechanism" in schools that alleviates the burden of teachers in classes with (potentially) less well-behaving children. The pattern described does not hold generally however, as the full sample correlation coefficient between class size and FBI was positive in both years (see Table 1). The reason might be that FBI and class size is positively correlated only in the segment somewhat more below the rule 2 threshold of 36. This is somewhat confirmed by a similar analysis performed using the thresholds stemming from the "soft rule", as in this case the estimated coefficients at the first two cutoffs have the opposite sign (see Table 15 in Appendix C). All in all, these results point to sorting but not between good schools with high enrollment and bad schools with low enrollment. Rather, it seems that schools with enrollment above a cutoff and thus with lower expected class size are attended by children with better backgrounds. All in all, I expect that the 2SLS estimates in the discontinuity sample will have a downward bias.

Table 3. The relationship of being above the cutoff points and family background index

	Year 2011 +/-5 sample	Year 2011 Full sample	Year 2013 +/-5 sample	Year 2013 Full sample
enrollment	.007***	.009	.007***	.009
	(.001)	(.000)	(.001)	(.000)
1[36<enrollment<42]	.140***	.190***	.052	.049
	(.052)	(.056)	(.051)	(.054)
1[72<enrollment<78]	.175***	.029	-.059	-.089
	(.064)	(.073)	(.060)	(.066)
1[90<enrollment<96]	.148	.074	.099	.059
	(.090)	(.106)	(.088)	(.098)
1[120<enrollment<126]	.004	.390**	-1.08***	-.807
	(.152)	(.153)	(.240)	(.236)
1[32<enrollment<42]		-.167***		-.018
		(.043)		(.040)
1[68<enrollment<78]		-.032		-.079*
		(.050)		(.047)
1[86<enrollment<96]		-.643***		-.496***
		(.087)		(.077)
1[116<enrollment<126]		-.134*		-.112
		(.078)		(.070)
R^2	.033	0.043	.0211	0.043
N	15753	64046	14645	56163

*Clustered standard errors are in parentheses. * p < 10%, ** p< 5% and ***p< 1%.The regressions in columns 2 and 4 were run in the +/-5 neighborhood of the cutoff points.*

The above reasoning shows that numerous problems may occur if one attempts to use expected class size as an instrument, and estimating the first differenced model can still be a better option resulting in a smaller bias. In order to eliminate one source of bias, in **other 2SLS specifications** I attempt to utilize the fact that pupils who were attending schools that were closed in the school year 2011/2012 had to enroll in another one. According to the law on public education[15] that was in force until September, 2012, if an institution ceases but has a legal successor, the students will be enrolled in that institution. If the parent does not want to enroll her child in the legal successor, or cessation is without a legal successor, the local government names the institution in which the parent can apply for her child's enrollment, and the assigned institution has to admit the child if it has enough places. As the law leaves the choice of the new institution to the local government, and it supposedly does not make decisions considering the abilities, achievement

[15] The Act LXXIX of 1993 on public education, paragraph 88.

and background of the children on a case by case basis, there is a potential for exogenous class size increases in school closures. Using the number of such pupils as an instrument for the change in class size in a 2SLS framework could partly eliminate the endogeneity stemming from the parents' decision to change school, although endogenous sorting can still occur. In order to alleviate the remaining sorting effects of this on the estimation, I restrict the sample to those students who attended schools that are in municipalities with population smaller than 5000, which decreases the chance that the students in the sample sort between schools. In other specifications I restrict the sample to those schools that 1) do not select among applicants based on their characteristics and 2) there is no similar school in the neighborhood according to the principal. I used three different instruments in the first stage regressions:

- the number of students in the class in 2013 who left their school due to its closure in the school 2011/2012;
- the number of those school leavers who attend a school in the same municipality they live, and their previous school was also there;
- the number of students whose school was situated in the same settlement (excluding county centers and the capital) in the previous period (2011) as the given non-changing student, but the institution was closed in the 2011/2012 school year (these results are not presented in the main text, only in the Appendix).

There are several potential problems with this approach. The first obvious issue is that some students may have otherwise left their institutions, a possibility that I could not check in this thesis.[16] Second, although closures were perhaps not based on student achievement or ability, they may be correlated with them through socio-economic status, which could be a more important

[16] A probit analysis may have the potential to estimate the propensities of students to leave based on the characteristics of students whose school was not closed, which could be a step towards identifying those who would otherwise have left. It is questionable, however, that based on the available data statistically significant coefficients can be obtained to construct reliable propensity scores.

factor that determined which schools closed. If these students were on average lower achieving, their enrollment in a new class could affect the achievement of others through peer effects other than increased probability of disruption due to class size increase. The third problem is nonrandom attrition, which I already mentioned during the discussion of the simple first differenced model. Fourth, the law and the restriction of the sample to similar schools do not guarantee the absence of sorting based on potential outcomes. Last but not least, the number of such school leavers is very small compared to the entire sample: the number of students leaving their schools due to closures in the 2013 sample is 3195, while the number of those who are for the second instrument is 2407, which constitute 4.5 and 3.4 percent of the 2013 sample, respectively. Furthermore, many of these students attended schools in towns larger than 5000, or in a legal successor of the original institution, which further reduces the variation in treatment that can be potentially explained by the instruments. The total number of non-changers (either due voluntarily or due to school closure) was 57554, and 5.59 percent of them were affected by the instrument.

Chapter 5: Results

5.1 First differenced specifications

First, I estimated a first differenced model with various controls, the results of which can be found in Table 4. These suggest that the characteristics of other students that are not controlled for in columns 1 and 2 were causing an omitted variables bias. Including other variables such as disadvantaged status and the ratio of disadvantaged students does not affect the results to large extent, therefore I omitted them from the final specification. It is worth noting that the change in the family background index had a remarkably high standard deviation of .394 in the sample (the standard deviation of the level variable is roughly equal to 1).[17] As it was emphasized above, even though time constant unobserved effects are removed from the estimated equation and I controlled for variables that largely capture factors underlying the endogeneity problem (e.g. family background), there are still numerous sources of omitted variables bias during the application of this method. The coefficient of class size is negative and significant for both dependent variables, as seen in the last two columns of Table 4, but their magnitudes are very small. Considering that the standard deviations of both tests scores are about 200, a five student increase in class size decreases achievement by about 0.011 standard deviations in mathematics and 0.016 standard deviations in reading comprehension. I attempted to reduce omitted variables bias adding more (first differenced) controls and including the lagged values of the explanatory variables as proxies for the initial values and past changes. As a result, the coefficients of class size did not change substantially; although in the equations for mathematics it remains only marginally significant on the 10% level (see Table 14 in Appendix C). The inference is based on cluster robust standard errors (as in the rest of the thesis), the consistency of which is supported by the large number of clusters (3593) (see e.g. Angrist and Pischke, 2009: 235). The per-student effect size (0.002σ to 0.003σ) estimated here are rather low in comparison to earlier results as well. Angrist and Lavy

[17] The reason is that it contains variables like having a computer and the number of books at home.

(1999) converted Krueger's (1999) results regarding the STAR experiment to a per-pupil effect size of about 0.048σ, while their own estimates ranged between 0.017σ and 0.071σ. As Krueger's paper is based on experimental data, it provides the strongest benchmark, and could also point to the presence of an upward bias (i.e. downward regarding the absolute value), in my results here. He also estimated that the cumulative effects of class size were small and not always statistically significant, and most of the effects could be attributed to the first year. This could be an explanation for the small estimates here as well, as class size effects may not be as large in 8^{th} grade, and the effect of being and staying in a small versus a large class is differenced out in the model estimated here.

If these estimates are biased, it is hard to tell its direction. Due to first differencing, time constant variables such as innate ability and "propensity to disrupt" were eliminated. These are the variables that the school management possibly responds to when it determines class sizes; therefore this source of bias is likely to have disappeared by first differencing. Past changes and the initial value, however, can still affect the current change in explanatory variables. In the case of class size, the initial (t=0) placement of children could matter the most. The equalization by school management and the efforts by parents of newcomers to place their children into smaller classes may imply an upward bias, as noted in the previous chapter. Another very realistic possibility in the Hungarian context, however, is within school segregation. If children with worse background are fewer in number and are usually placed in the same, segregated class which are therefore smaller than the average, newcomers will be placed, and As the two effects can be present simultaneously, it is not possible to tell how true class size effects differ from the estimates in the first differenced model.

Table 4. Results of the first differenced regressions

	(1) D. Mathematics	(2) D. Reading	(4) D. Mathematics	(5) D. Reading
D. class size	-.287	-1.119***	-.446**	-.626***
	(.483)	(.381)	(.212)	(.202)
D. enrollment	-.309	.028	-.348***	.090
	(.231)	(.168)	(.115)	(.099)
D. family background index	5.801***	5.957***	3.668*	3.607**
	(2.193)	(1.866)	(1.630)	(1.501)
D. average family background index of other students in the class			-43.439***	-32.940***
			(2.976)	(2.959)
D. average test score of other students in the class			.820***	.695***
			(.009)	(.013)
N	38585	38573	38570	38558
R²	0.0006	0.0012	0.2980	0.1712

Clustered standard errors are in parentheses. $* p < 0.1$, $** p < 0.05$ *and* $*** p < 0.01$.

5.2 Class size function as an instrument

Table 5 shows the results of the first stage regressions of the 2SLS specification using rule 2. Adding enrollment and its second order polynomial as a control halves the coefficient of expected class size, but adding the family background index to the explanatory variables does not affect it. The coefficients are highly significant, which implies that expected class size can be considered a strong instrument. The second stage results presented in Table 6 are somewhat surprising. Without any controls, the effect of class size is estimated to be positive and statistically significant and around ten times the magnitude of the coefficients estimated by first differencing. Including enrollment decreases the point estimates with regard to both years, but for 2013 they still remain statistically significant. Using the second order polynomial of enrollment results in radical changes: the coefficients of class size change sign and become statistically significant only in regarding year 2011. After adding FBI as an explanatory variable, the coefficient of class size loses most of its statistical significance. The pattern of changing sign suggests that expected class size is not a valid instrument, at least for the full sample, and changes in C_f are mostly led by changes in enrollment. As enrollment is positively correlated with FBI[18] and possibly other background variables that are not necessarily included in the index (e.g. because larger schools are usually found in wealthier regional centers), these omitted variables may affect the coefficient of class size as well. Moreover, if the functional form is not adequately defined, the coefficient of class size possibly takes up some of the nonlinear effects of enrollment. To account for these effects, I also ran regressions including piecewise linear trends with the same slopes as C_f besides the controls in Table 6, but obtained similar results (the 2011 results are in Table 16 in Appendix C). Full sample results of applying rule 1 can be found in Tables 19-20 in Appendix C. The effect of class size are shown to be positive and statistically significant until FBI is included as a regressor. I argued above that this rule is not a valid instrument even around the cutoffs

Table 5. Full sample first stage regressions. The dependent variable is class size, the class size cap is defined by rule 2.

	Year 2011				Year 2013			
Expected class size (rule 2)	.444*** (.013)	.285*** (.016)	.210*** (.018)	.214*** (.018)	.355*** (.012)	.205*** (.015)	.124*** (.017)	.132*** (.018)
enrollment		.072*** (.004)	.181*** (.010)	.248*** (.009)		.075*** (.004)	.207*** (.012)	.198*** (.012)
(enrollment²) /100			-.085*** (.007)	-.126*** (.009)			-.109*** (.009)	-.108*** (.009)
Family background index				.604*** (.042)				.492*** (.046)
N	80287	80287	80287	64046	70661	70661	70661	56163
R²	.246	.340	.361	.376	.184	.274	.303	.313

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01.*

[18] Their correlation was 0.195 in 2011 and 0.197 in 2013.

Table 6. Full sample 2SLS regression results.

	Year 2011		Year 2013		Year 2011		Year 2013	
	Mathematics	**Reading**	**Mathematics**	**Reading**	**Mathematics**	**Reading**	**Mathematics**	**Reading**
class size	3.592***	4.730***	7.371***	6.869***	.060	.755	3.664**	3.074**
	(.687)	(.588)	(.816)	(.756)	(1.29)	(1.121)	(1.717)	(1.567)
enrollment					.712**	.801**	.659***	.674***
					(.164)	(.144)	(.201)	(.182)
N	80287	80287	70661	70661	80287	80287	70661	70661

	Year 2011		Year 2013		Year 2011		Year 2013	
	Mathematics	**Reading**	**Mathematics**	**Reading**	**Mathematics**	**Reading**	**Mathematics**	**Reading**
class size	-5.836***	-4.655***	-3.382	-3.345	-3.193*	-1.813	-1.848	-1.370
	(2.099)	(1.845)	(3.311)	(3.052)	(1.678)	(1.233)	(2.332)	(1.910)
enrollment	3.588***	3.440***	3.545***	3.304***	1.088**	.795**	1.088*	.698
	(.635)	(.559)	(.933)	(.872)	(.464)	(.338)	(.631)	(.518)
enrollment^2/100	-1.91***	-1.752***	-1.960***	-1.786***	-.660**	-.437**	-.659*	-.410
	(.379)	(.333)	(.559)	(.530)	(.267)	(.193)	(.379)	(.312)
Family background index					87.517***	92.065***	94.908***	99.541***
					(1.670)	(1.295)	(1.712)	(1.475)
N	80287	80287	70661	70661	64046	64046	56163	56163

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01.*

As the literature using the method emphasizes, the (fuzzy) regression discontinuity approach is only reliable if it is applied around the discontinuity points, and the underlying population is similar at the two sides of them. Although I find it highly plausible that the latter condition is not met (see above) Table 7 shows the results of performing the first stage regression in the +/-5 neighborhood of discontinuity did not result in a statistically significant coefficient on class size except in the case when expected class size was the only regressor. This is the result of the pattern observed in the previous chapter: schools mostly add one more class well below reaching the class size cap due to "historical" reasons. For this reason, C_f is a weak instrument and cannot be used for inference. Nevertheless, the second stage results for 2011 are shown in Table 17 in Appendix C. The estimated coefficients are statistically not different from 0, which is possibly the result of the weak correlation between C_f and class size (see e.g. Wooldridge, 2002: 470). The large negative point estimates here possibly stem from the negative correlation between the error term and the instrument, and from the observation that the availability of FBI is not independent from class size and achievement. First stage regressions on the +/-1 discontinuity sample are also presented in the Appendix; none of the specifications yielded statistically significant coefficient estimates for the class size function in this case.

Table 7. First stage regressions for the 2SLS specification in the +/-5 neighborhood of discontinuity points

	Class size, 2011	Class size, 2013	Class size, 2011	Class size, 2013	Class size, 2011	Class size, 2013
Expected class size	.044** (.021)	.001 (.021)	.010 (.019)	-.026 (.018)	.006 (.019)	-.029 (.019)
enrollment			.113 (.006)	.119 (.007)	.196*** (.031)	.254*** (.038)
(enrollment^2)/100					-.062*** (.022)	-.110*** (.030)
N	19705	18604	19705	18604	19705	3909
R^2	0.0036	0.000	0.353	0.308	0.360	0.319

*Clustered standard errors are in parentheses. The discontinuity sample is +/-5 around the enrollment values where the expected class size function has a discrete jump, * p < 0.1, ** p< 0.05 and ***p< 0.01.*

First stage regressions using rule 1 in the neighborhood of the discontinuity points show a much stronger relationship between C_f and class size in the full sample, and the coefficients are statistically significantly different from zero even in the neighborhood of the discontinuity points (Tables 21 and 24). The second stage estimates for the coefficient of class size are positive and significant only in the +/-5 discontinuity sample of 2011 (Tables 22, 23 and 25), but they also lose their significance if family background index is included in the regressions or the sample is restricted as before. Due to the "soft" nature of rule 1, it is most likely not randomly applied, and it is reasonable to assume positive correlation between the error term and expected class size around the discontinuity points. Therefore, the positive bias that can be assumed in the case of simple OLS remains in the second stage as well.

5.3 Number of school changers due to closures as an instrument

It is worth noting beforehand that due to the notable number of possible endogeneity problems and the statistical insignificance of the coefficient estimates in every specification, I present only some of the results in this section. I also attempted applying other combinations of sample restrictions, but none of them resulted in statistically significant coefficients. Table 8 shows the results of first stage regressions with the first two instruments: 1) the number of school changers due to school closures and 2) the number of those among them who stayed in the same municipality. In columns (1) and (2) the sample is restricted to settlements with population smaller than 5000; whereas columns (3) and (4) considers only those schools that do not select students based on the applicants characteristics and where there is no similar school in the neighborhood according to the principal. The only relevant instrument is the first one, and its coefficient is significantly different from zero only when the second restriction is applied. These regressions also included the main controls; the regressions with fewer controls resulted in essentially the same coefficients (see Table 26 in Appendix C).

Table 8. First stage regressions with number of new students from closed schools as instrument.

	Sample based on the population of the settlement		Sample based on the questionnaire for principals	
	(1) D. class size	(2) D. class size	(3) D. class size	(4) D. class size
No. of school changers due to closures	0.647		0.944**	
	(0.432)		(0.427)	
D.FBI	-0.0395	-0.0388	-0.0324	-0.0215
	(0.0728)	(0.0728)	(0.0834)	(0.0846)
D.enrollment	0.347***	0.350***	0.128***	0.130***
	(0.0549)	(0.0546)	(0.0307)	(0.0309)
D.mean FBI of other children in the class	0.153	0.135	0.363	0.320
	(0.418)	(0.418)	(0.532)	(0.535)
No. of school changers due to closures who stayed		2.598		0.709
		(2.136)		(0.636)
N	11055	11055	7791	7791
R^2	0.143	0.146	0.075	0.069

*Clustered standard errors in parentheses, * p < 0.1, ** p< 0.05 and ***p< 0.01. Note: in columns (1) and (2) only settlements with population lower than 5000 were used. In columns (3) and (4), the sample is restricted to those schools that 1) do not select among applicants and 2) there is no similar school in the neighborhood according to the principal.*

The coefficients resulting from the second stage regressions are large but their difference from zero statistically not significant (see Table 9). The explanation for these partly lies in the fact that the sample is restricted to great extent and there are only a relatively few "treated" individuals (those who did not change class, yet their class size changed), that is, the variation in the explanatory variables is rather small. The multicollinearity issue is coupled with high correlation between the clusters and estimated by 2SLS, which are also known to increase standard errors (although the R^2 is not very high in the first stage regressions, see e.g. Wooldridge 2002: 479). The coefficients are rather larger large, exceeding 0.1σ.[19] The large negative coefficients can be the results of a downward bias due to the correlation between the change in class size explained by the hypothetical instrument and unobserved factors, e.g. the involuntary school changers get into historically lower quality schools due to their initial lower achievement and worse family background on average.[20]An important omitted variable can be the "general economic situation" of the region which could impact the change in scores and the number of school closures as well. Adding FBI and other explanatory variables further decreases the sample due to the unavailability of this variable for some students, and the unavailability of the FBI for some students causes downward bias (see the chapter on data and variables).

Another possibility is that effects are heterogeneous, and lower achieving students lose more by increasing their class size. Krueger (1999) estimated heterogeneous effects in the case of the STAR experiment; students on free lunch and black pupils were shown to gain more from class size reduction. The heterogeneity of effects shown in the study was not that large, however, and the estimated coefficients had similar magnitudes like the ones for the full sample, which were less than the half of the estimates here. Thus, the results presented in this section should be considered

[19] The standard deviation of scores in this subsample was 195.47 for mathematics and 190.87 for reading comprehension in 2013.
[20] The FBI of these school changers in this restricted sample was -.331, which is well below the full sample average, but the subjects of the estimation (non-changers) had an even lower average value of -.407 in the subsample.

inconclusive regarding the answer to the research question. Using the number of school leavers due to school closures in the settlement as an instrument for the change in class size did not yield significant results either, and the coefficients of instruments in the first stage regressions were significant only when all towns were considered except for Budapest and county centers. The 2SLS estimates did not result in statistically non-zero coefficients in this case either; these estimates can be found in Tables 28 and 29 in Appendix C.

Table 9. 2SLS regressions with the number of new students from closed schools as instrument.

	(1) D.Mathematics	(2) D.Reading	(3) D.Mathematics	(4) D.Reading
D.class size	-21.60	-10.42	-36.14	-25.72
	(13.18)	(11.33)	(26.38)	(23.39)
D.enrollment			3.901	3.445
			(3.679)	(3.148)
D.FBI			8.339	6.607
			(6.032)	(4.382)
D.mean FBI of other pupils in the class			49.58	18.57
			(27.94)	(19.10)
N	10804	10796	7787	7781

Clustered standard errors in parentheses. None of the presented coefficients were statistically significant. Note: the sample is restricted to those schools that 1) do not select among applicants and 2) there is no similar school in the neighborhood according to the principal.

Conclusions

The aim of this thesis was to estimate the effect of class size on student achievement in Hungary. The subject of the empirical analysis was the scores achieved by a cohort of students in the 6th and 8th grades on the National Assessment of Basic Competences, a yearly administered standardized test in mathematics and reading comprehension. The majority of credible empirical analyses based on experimental and quasi-experimental found evidence that the effect is negative and the per-student effect size is somewhere around 0.05 standard deviations. These effects, however, were observed in the lower grades of primary schools.

Although I used more than one potential instrument in order to estimate the relationship between the two variables, I also argued that none of these variables are credibly exogenous to the potential outcomes of students. In the case of class size function, I found it plausible that various types of schools and students did not apply the maximum rule(s) uniformly. In addition, the relevance of the instrument in explaining variation in class size was especially low in the subsample where the instrument could estimate causal effects most reliably. The ostensible reason is that in the 6th and 8th the number of classes is primarily based on the class size rule of previous grades which defines a considerably lower maximum, and schools usually do not admit so many new students that they have to launch new classes. Therefore, average class size is much lower than would be expected based on enrollment. Thus, I argue that the instrument is not independent of unobserved variables regarding the full sample (a statement that cannot be entirely investigated empirically), and not relevant around the class size rule (which is apparent from the first stage regressions). The instruments based on the number of students that had to leave their previous school due to its closure were only relevant in some of the subsamples, and there are reasons why they may not be valid; most importantly, the legal setting and the restriction of the sample to similar schools do not guarantee the absence of sorting based on potential outcomes. Last but not least, the number of such school leavers is very small compared to the entire sample, and the number of treated individuals is thus also very low, which makes estimating statistically significant coefficients more difficult.

Due to the above reasons, I find the estimates based on the first differenced model the most reliable from the ones presented in this thesis. Although there are numerous reasons why these results may be biased, and the estimates are about ten times smaller than to the most credible ones in the empirical literature, these are still the closest to them due to the statistical insignificance of the 2SLS results. It is also worth noting that the study by Krueger (1999) based on the Tennessee STAR experiment found very small cumulative effects; and it may indeed be the case that class size has very little or no effect in the 6^{th} and 8^{th} grade.

The further investigation of the topic using Hungarian data could exploit between cohort changes regarding the population of students and the number of classes in the same schools, similarly to the analysis of Hoxby (2000). In addition, the NABC data is in large part based on a questionnaire filled in by principals and students. It would be preferable to use more data from other sources by which the economic and demographic environment of the schools could be controlled for.

Appendix A

If $p = p(q)$ and $p'(q) \geq 0$, the comparative statics for q when profit is maximized:

$$\frac{\partial m}{\partial q}|_{FOC} = \frac{v\frac{Z^2}{m^2}\ p^{\frac{Z}{m}}\ln(p)\ +\ v\frac{Z^3}{m^3}\ q\ p^{\frac{Z}{m}-1}\ln(p)\ p^{'}(q) +\ v\frac{Z^2}{m^2}\ q\ p^{\frac{Z}{m}-1}p^{'}(q)}{\frac{\partial^2\pi}{\partial m^2}}$$

This is negative if the numerator is positive, i.e.:

$$v\ \frac{Z^2}{m^2}p^{\frac{Z}{m}-1}\ [p\ \ ln(p) + \frac{Z}{m}\ ln(p)\ q\ p'(q) +\ q\ p'(q)] > 0$$

The term in the square brackets is not positive for sufficiently low p and high Z/m. E.g. if $p = 0.95$ and $Z/m = 25$, it is negative, thus an increase in q in this case leads to an increase in class size. This value of p is very low, however, and with a class size of 25, it leads to actual teaching only 27.7 percent of the time. At a more reasonable value of 0.98, class size (Z/m) would have to be about 50 for the expression to be negative. On the other hand, the derivative with respect to p is:

$$\frac{\partial m}{\partial p}|_{FOC} = \frac{v\frac{Z^2}{m^2}\ p^{\frac{Z}{m}-1}[\frac{Z}{m}\ q\ln(p) + p\ln(p) + q]}{\frac{\partial^2\pi}{\partial m^2}}$$

This is positive only if the expression in square brackets is negative, but for realistic values of p, Z/m should be very large for this to be true. E.g. if $p = 0.99$, Z/m has to be close to 100 so that $-(Z/m)\ q\ ln(p) = q$, as $p\ ln(p)$ is a very small negative number.

Appendix B

Table 10. Education at primary schools in the past decade

School year	Schools, school sites	Classrooms	Teachers
2001/2002	3 852	43 195	90 294
2002/2003	3 793	42 603	89 035
2003/2004	3 748	42 051	89 784
2004/2005	3 690	41 581	87 116
2005/2006	3 614	40 980	85 469
2006/2007	3 591	40 513	83 606
2007/2008	3 418	38 784	78 073
2008/2009	3 363	37 952	75 606
2009/2010	3 343	37 463	74 241
2010/2011	3 306	37 045	73 565
2011/2012	3 252	36 338	72 501
2012/2013	3 251	35 960	72 048

School year	Students		Classes in full time education	Students per class	Students per teacher
	total	in 5th to 8th (10th) grades			
2001/2002	947 037	466 660	47 682	19,8	10,5
2002/2003	933 171	466 621	46 539	20,0	10,4
2003/2004	912 959	463 159	45 774	19,9	10,1
2004/2005	890 551	457 224	44 883	19,8	10,2
2005/2006	861 858	443 457	43 475	19,8	10,1
2006/2007	831 262	429 693	42 266	19,6	9,9
2007/2008	811 405	414 914	39 900	20,3	10,4
2008/2009	790 722	399 129	38 799	20,3	10,4
2009/2010	775 741	385 737	38 140	20,3	10,4
2010/2011	758 566	369 611	37 475	20,2	10,3
2011/2012	749 865	362 767	36 810	20,3	10,3
2012/2013	745 058	357 696	36 672	20,3	10,3

Source: the website of the Hungarian Central Statistical Office (KSH)

Table 11. Statistics of regular elementary schools

Year	Variable	Mean	Standard deviation	Minimum	Maximum	Number of observations
2011	**Reading score**	1459.883	189.456	811.1749	2166.955	80287
	Math score	1481.114	201.434	849.0269	2264.114	80287
	FB index	-.0658637	1.011246	-3.169531	1.989724	64046
	Class size	22.56598	5.18609	1	36	80287
2013	**Reading score**	1541.25	194.3349	746.96	2185.341	70661
	Math score	1606.363	198.8985	901.224	2293.656	70661
	FB index	-.1218185	.9938184	-3.190755	1.897206	56163
	Class size	21.39561	5.094301	1	37	70661

The years apply for members of the same cohort, who were 6^{th} graders in 2011.

Table 12. Statistics of eight-grade high schools

Year	Variable	Mean	Standard deviation	Minimum	Maximum	Number of observations
2011	**Reading score**	1616.292	157.7588	1037.695	2093.497	3652
	Math score	1624.168	176.3144	934.1718	2143.591	3652
	FB index	.8680741	.7489315	-2.124888	1.989724	2988
2013	**Reading score**	1707.17	152.1582	995.3761	2160.455	3464
	Math score	1766.116	169.2295	1100.886	2293.656	3464
	FB index	.8483967	.714304	-2.371507	1.897206	2781

The years apply for members of the same cohort, who were 6^{th} graders in 2011.

Table 13. Statistics of six-grade high schools

Year	Variable	Mean	Standard deviation	Minimum	Maximum	Number of observations
2013	**Reading score**	1697.397	159.6716	978.3721	2199.417	4765
	Math score	1756.628	180.0954	991.0917	2258.873	4765
	FB index	.8565887	.7155919	-1.897206	1.897206	3974

The years apply for members of the same cohort, who were 6^{th} graders in 2011.

Figure 4. Distribution of class size in the sample

Figure 5. Scatterplot of class size and reading comprehension

Figure 6. Scatterplot of family background index and scores in mathematics

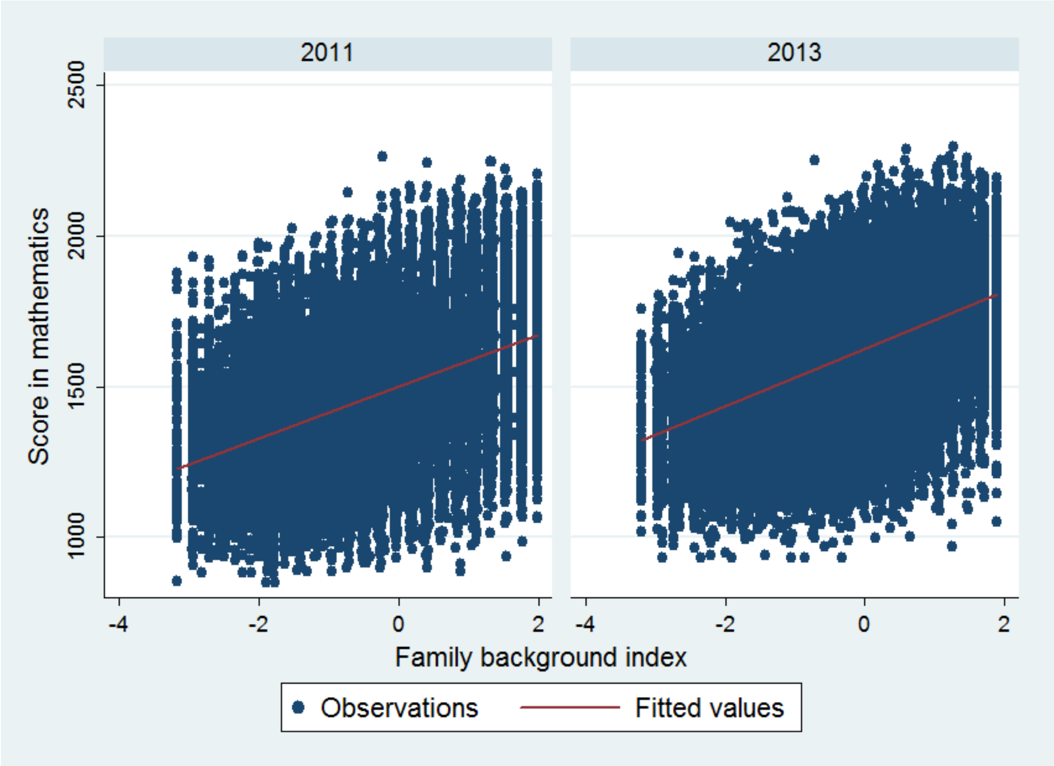


Figure 7. Scatterplot of family background index and scores in reading

Appendix C

Table 14. Results of first differenced regressions with additional controls (continued on page 60)

	(1) D.Mathematics	(2) D.Reading	(3) D.Mathematics	(4) D.Reading
D.class size	-0.353	-0.599***	-0.364*	-0.605***
	(0.215)	(0.209)	(0.214)	(0.207)
D.FBI	6.391***	8.890***	6.515***	8.979***
	(1.684)	(1.553)	(1.683)	(1.553)
L.FBI	6.855***	12.23***	6.863***	12.23***
	(0.901)	(0.851)	(0.900)	(0.851)
L.enrollment	-0.0427	-0.0147	-0.0461	-0.0167
	(0.0284)	(0.0269)	(0.0282)	(0.0269)
L. average math score of other students in the class	-0.0529***		-0.0518***	
	(0.00743)		(0.00743)	
L.average FBI of other students in the class	10.83***	0.939	10.59***	0.763
	(1.522)	(1.711)	(1.522)	(1.710)
L.class size	0.0174	-0.0385	0.0226	-0.0312
	(0.142)	(0.143)	(0.141)	(0.143)
D.average math score of other students in the class	0.775***		0.777***	
	(0.0110)		(0.0109)	

Table 14 (continued). Results of first differenced regressions with additional controls

	(1) D.Mathematics	(2) D.Reading	(3) D.Mathematics	(4) D.Reading
D. average FBI of other students in the class	-37.31***	-32.25***	-34.86***	-30.43***
	(2.998)	(3.053)	(2.999)	(3.044)
D.enrollment	-0.315*	-0.0438	-0.331***	-0.0557
	(0.124)	(0.108)	(0.123)	(0.107)
D.ratio of disadvantaged in the class			27.30***	18.69**
			(8.350)	(7.952)
D.ratio of ILB students in the class			28.51***	29.32***
			(9.832)	(9.187)
L. average reading score of other students in the class		-0.0279**		-0.0274***
		(0.00951)		(0.00951)
D. average reading score of other students in the class		0.674***		0.676***
		(0.0135)		(0.0135)
N	38570	38558	38570	38558

*Clustered standard errors are in parentheses. * $p < 0.1$, ** $p< 0.05$ and ***$p< 0.01$. Note: ILB stands for students having problems with integration, behavior and learning.*

Table 15. Regression of FBI on being enrolled around the discontinuity points using rule 1.

	Year 2011 +/-5 sample	Year 2011 Full sample	Year 2013 +/-5 sample	Year 2013 Full sample
enrollment	.004***	.008***	.005***	. 009***
	(.001)	(.000)	(.001)	(.000)
1[30<enrollment<36]	-.296***	-.146**	-.195***	-.021
	(.055)	(.061)	(.052)	(.058)
1[60<enrollment<66]	.079**	-.148***	.086**	-.096*
	(.039)	(.057)	(.039)	(.056)
1[90<enrollment<96]	.085	.076	.072	.062
	(.072)	(.107)	(.072)	(.097)
1[120<enrollment<126]	-.244	.391**	-.276*	-.805***
	(.157)	(.154)	(.157)	(.235)
1[25<enrollment<36]		.018		-.005
		(.045)		(.043)
1[55<enrollment<66]		.189***		.048
		(.039)		(.041)
1[85<enrollment<96]		-.595***		-.437***
		(.086)		(.076)
1[115<enrollment<126]		-.097		-.073
		(.077)		(.069)
R^2	.032	.045	.028	.042
N	28736	64046	28784	56163

Clustered standard errors are in parentheses. $*p < 0.1$, $**p < 0.05$ *and* $***p < 0.01$

Table 16. 2SLS regressions with piecewise linear trends added

Rule 1 ("soft rule")	Math, 2011	Reading, 2011	Rule 2 ("hard rule")	Math, 2011	Reading, 2011
class size	1.506 (.932)	1.271* (.707)	**class size**	-3.235** (1.444)	-1.556 (1.065)
enrollment	2.529 *** (.939)	1.742 ** (.707)	**enrollment**	2.470 (1.539)	1.841 (1.124)
enrollment^2/100	-9.562*** (2.683)	-5.803 *** (2.109)	**e^2/100**	-3.683 (2.886)	-2.737 (2.088)
Family background index	84.665*** (1.390)	90.241*** (1.108)	**Family background index**	87.212*** (1.521)	91.721*** (1.190)
((e - 30)/2)*1[30<e]	12.223*** (3.497)	6.944** (2.762)	**((e - 36)/2)*1[36<e]**	3.656 (3.481)	2.542 (2.513)
((e - 60)/3)*1[60<e]	16.773*** (4.988)	10.807*** (3.887)	**((e - 72)/3)*1[72<e]**	5.953 (5.979)	5.991 (4.319)
((e - 90)/2)*1[90<e]	21.896*** (7.150)	12.499** (5.614)	**((e - 90)/2)*1[90<e]**	4.001 (5.279)	.858 (3.725)
((e - 120)/2)*1[120<e]	12.657 (13.526)	8.198 (12.111)	**((e - 120)/2)*1[120<e]**	-5.000 (15.593)	-.040 (13.154)
((e - 150)/2)*1[150<e]	83.286*** (25.193)	49.131* (25.189)	**((e - 150)/2)*1[150<e]**	60.944** (28.332)	37.012 (26.954)
N	**64046**	**64046**	**N**	**64046**	**64046**

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01. Note: "e" stands for enrollment.*

Table 17. 2SLS regressions for 2011 on the +/-5 discontinuity sample, rule 2

	Discontinuity sample: +/-5							
	Math	**Reading**	**Math**	**Reading**	**Math**	**Reading**	**Math**	**Reading**
Class size	-23.705 (19.015)	-22.711 (17.634)	-123.84 (242.649)	-124.616 (244.114)	-223.413 (739.079)	-225.562 (748.213)	-18.459 (37.938)	-22.657 (35.947)
enrollment			14.708 (27.854)	14.967 (28.038)	46.388 (148.819)	47.085 (150.714)	3.745 (7.199)	4.624* (6.981)
(enrollment^2) /100					-15.240 (48.257)	-15.451 (48.885)	-1.290 (2.378)	-1.534* (2.335)
Family background index							99.323* (22.024)	104.904* (20.637)
N	19705	19705	19705	19705	19705	19705	15753	15753

*Clustered standard errors are in parentheses. * p < 0.01.*

Table 18. First stage regressions for the 2SLS specification in the immediate neighborhood of discontinuity points (+/-1), rule 2

	Class size, 2011	Class size, 2013	Class size, 2011	Class size, 2013	Class size, 2011	Class size, 2013
Expected class size	.018 (.0363)	.020 (.042)	.037 (.028)	-.024 (.037)	.033 (.029)	-.027 (.038)
enrollment			.135*** (.013)	.145*** (.014)	.249 (.145)	.282** (.141)
(enrollment^2)/100					-.096 (.127)	-.117 (.122)
N	4279	3909	4279	3909	4279	3909
R^2	0.0010	0.0011	0.437	0.412	0.439	0.416

*Clustered standard errors are in parentheses. The discontinuity sample is +/-1 around the enrollment values where the expected class size function has a discrete jump. * p < 0.1, ** p< 0.05 and ***p< 0.01.*

Table 19. Full sample first stage regressions, rule 1

	Year 2011				Year 2013			
Expected class size (rule 1)	.756*** (.014)	.651*** (.019)	.600*** (.021)	.592*** (.022)	.687*** (.014)	.589*** (.020)	.540*** (.023)	.530*** (.023)
enrollment		.033*** (.003)	.105*** (.008)	.100*** (.008)		. 033*** (.004)	.107*** (.009)	.105*** (.010)
(enrollment^2)/100			-.057*** (.006)	-.055*** (.006)			-.062*** (.007)	-.063*** (.008)
Family background index				.486*** (.042)				.428*** (.042)
N	80287	80287	80287	64046	70661	70661	70661	56163
R^2	0.471	.488	.499	.508	.408	.423	.433	0.436

Clustered standard errors are in parentheses. All coefficients are significant on the 1% level.

Table 20. Full sample 2SLS regressions, rule 1

	Year 2011		Year 2013		Year 2011		Year 2013	
	Math	**Reading**	**Math**	**Reading**	**Math**	**Reading**	**Math**	**Reading**
class size	1.929** (.874)	2.489*** (.764)	1.934** (.974)	2.210** (.893)	.454 (.774)	.655 (.574)	.459 (.799)	.525 (.667)
enrollment	1.594*** (.332)	1.606*** (.289)	2.163*** (.373)	3.304*** (1.861)	.183 (.273)	.182 (.201)	.494* (.292)	.210 (.241)
enrollment² /100	-.912*** (.229)	-.835*** (.199)	-1.204*** (.274)	-.994*** (.261)	-.199 (.181)	-.126 (.133)	-.323 (.209)	-.135 (.171)
Family background index					85.404*** (1.368)	90.634*** (1.090)	93.803*** (1.289)	98.633*** (1.179)
N	80287	80287	70661	70661	64046	64046	56163	56163

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01*

Table 21. First stage regressions using the +/-5 sample, rule 1

	Class size, 2011	Class size, 2013	Class size, 2011	Class size, 2013	Class size, 2011	Class size, 2013
Expected class size	.370*** (.036)	.376*** (.035)	.331*** (.037)	.341*** (.037)	.331*** (.037)	.341*** (.037)
enrollment			.032*** (.006)	.038*** (.008)	.035 (.027)	.001 (.033)
(enrollment²)/100					-.002 (.019)	.032 (.026)
N	23470	18470	23470	18470	23470	18470
R^2	0.1036	0.1193	0.1234	0.137	0.1234	0.1383

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01.*

Table 22. 2SLS regressions on the +/-5 sample, rule 1, year 2011

	Discontinuity sample: +/-5							
	Math	**Reading**	**Math**	**Reading**	**Math**	**Reading**	**Math**	**Reading**
Class size	6.628*** (1.884)	6.505*** (1.638)	5.155** (2.150)	4.681** (1.904)	5.075** (2.150)	4.589** (1.905)	2.928 (1.956)	1.914 (1.509)
enrollment			.461** (.186)	.571*** (.165)	1.475** (.670)	1.746*** (.582)	-.502 (.571)	-.312 (.403)
(enrollment²) /100					-.791 (.515)	-.916*** (.443)	.335 (.439)	-.249 (.300)
Family background index							85.462*** (2.960)	92.274*** (2.344)
N	23470	23470	23470	23470	23470	23470	18667	18667

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01.*

Table 23. 2SLS regressions on the +/-5 sample, rule 1, year 2013, results for mathematics

	Mathematics, discontinuity sample +/-5			
Class size	2.695 (1.885)	.434 (2.168)	.478 (2.156)	-.799 (1.768)
enrollment		.940** (.257)	4.164** (.970)	2.087** (.780)
(enrollment^2)/100			-2.784** (.868)	-1.499* (.699)
Family background index				98.569** (2.518)
N	18470	18470	18470	14765

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01.*

Table 24. First stage regressions for the 2SLS specification in the immediate neighborhood of discontinuity points (+/-1), rule 1

	Class size, 2011	**Class size, 2013**	**Class size, 2011**	**Class size, 2013**	**Class size, 2011**	**Class size, 2013**
Expected class size	.199*** (.076)	.124 (.082)	.169** (.078)	.064 (.088)	.143* (.081)	.066 (.087)
enrollment			.047** (.021)	.069*** (.025)	.229* (.118)	-.038* (.120)
(enrollment^2)/100					-.160 (.100)	-.096 (.100)
N	4230	3336	4279	3909	4230	3336
R^2	.034	.015	0.055	0.065	0.065	0.069

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01.*

Table 25. 2SLS regressions in the immediate neighborhood of discontinuity points (+/-1), rule 1

	Year 2011				Year 2013			
Class size	3.193	-.9046	.716	2.928	-4.087	-30.959	-31.140	-11.254
	(6.953)	(8.751)	(10.265)	(1.956)	(12.361)	(52.982)	(51.496)	(18.602)
enrollment		1.291	-.745	-.502		3.873	4.485	3.649
		(.685)	(3.996)	(.571)		(4.031)	(5.403)	(2.43)
(enrollment²)/100			1.734	.335			-.536	-2.298
			(3.108)	(.439)			(6.685)	(2.400)
Family background index				85.462***				108.577***
				(2.960)				(20.979)
N	4230	4230	4230	3278	3336	3336	3336	2664

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01*

Table 26. First stage regressions with the number of new students from closed schools as instrument.

	(1) D.class size	(2) D.class size	(3) D.class size	(4) D.class size
No. of school changers due to closures	1.529***		0.933**	
	(0.490)		(0.427)	
No. of school changers due to closures who stayed		2.049*		0.717
		(0.916)		(0.635)
D.FBI			-0.0440	-0.0342
			(0.0907)	(0.0919)
D.enrollment			0.130***	0.132***
			(0.0306)	(0.0308)
N	11143	11143	7796	7796
***R*²**	0.017	0.007	0.074	0.069

*Clustered standard errors in parentheses, * p < 0.1, ** p< 0.05 and ***p< 0.01. Note: the sample is restricted to those schools that 1) do not select among applicants and 2) there is no similar school in the neighborhood according to the principal.*

Table 27. Regressions of class size change on the number of school leavers due to school closures in towns with population less than 5000 people

	(1) D.class size	(2) D.class size	(3) D.class size	(4) D.class size
No. of school changers due to closures	0.940*		0.646	
	(0.472)		(0.432)	
No. of school changers due to closures who stayed		2.588		2.603
		(1.844)		(2.138)
D.FBI			-0.0491	-0.0488
			(0.0802)	(0.0803)
D.enrollment			0.347***	0.350***
			(0.0549)	(0.0546)
N	15934	15934	11062	11062
R^2	0.005	0.005	0.142	0.145

*Clustered standard errors in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01*

Table 28. First stage regressions, the lagged number of school leavers due to school closures in the area is used as instrument

	D. class size	
Lagged number of school leavers due to school closures by postal code	.016*** (.004)	
Lagged number of school leavers due to school closures by municipality		.015*** (.004)
N	39400	39400

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01. Note: county seats and Budapest were excluded from the sample due to their sizes and central role.*

Table 29. 2SLS regressions, the lagged number of school leavers due to school closures in the area is used as instrument

	Mathematics				Reading			
	Area defined by postal code		Area defined as municipality		Area defined by postal code		Area defined as municipality	
D.class size	-5.886 (7.183)	-4.445 (8.647)	-5.246 (7.492)	-3.567 (9.009)	-.909 (4.112)	-3.135 (5.579)	-.669 (4.322)	-3.298 (5.891)
D.enrollment		.326 (1.494)		.1775 (1.554)		.375 (.941)		.402 (.992)
D.Family background index		8.314*** (2.655)		8.291*** (2.659)		6.024*** (2.284)		6.027*** (2.284)
N	38169	27144	38169	27144	38164	27136	38164	27136

*Clustered standard errors are in parentheses. * p < 0.1, ** p< 0.05 and ***p< 0.01. Note: county seats and Budapest were excluded from the sample due to their sizes and central role.*

References

Act LXXIX of 1993 on public education. Accessed at http://www.okm.gov.hu/letolt/kozokt/kozokt_tv_070823.pdf. Last retrieval June 1, 2015.

Angrist, Joshua D., and Guido Imbens. 1995. "Average Causal Response in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90 (430), 431–442.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Angrist, Joshua D., and Victor Lavy. 1999**. "**Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *The Quarterly Journal of Economics* 114 (2), 533-575.

Auxné Bánfi Ilona, Balázsi Ildikó, Lak Ágnes Rozina and Szabó Vilmos. 2012. **"**Országos Kompetenciamérés 2011, országos jelentés." Oktatási Hivatal, 2012.

Balázsi Ildikó, Lak Ágnes Rozina, Szabó Vilmos, Szabó Lívia Dóra, Vadász Csaba. 2014. **"**Országos Kompetenciamérés 2013, országos jelentés." Oktatási Hivatal, 2013.

Balázsi Ildikó and Zempléni András. 2004. "A hozottérték-index és a hozzáadott pedagógiai érték számítása a 2003-as kompetenciamérésben." *Új Pedagógiai Szemle,* 8 (10), 36–50.

Bohrnstedt, George W., and Brian M. Stecher (Eds.). **2002**." What we have learned about class size reduction." Sacramento, CA: California Department of Education. Accessed at http://www.classize.org/techreport/CSRYear4_final.pdf. Last retrieval: June 2, 2015.

Boozer, Michael, and Cecilia Rouse. 1995. "Intraschool Variation in Class Size: Patterns and Implications," NBER Working Paper No. 5144

Browning, Martin, and Eskil Heinesen. 2007. "Class Size, teacher hours and educational attainment." *The Scandinavian Journal of Economics*, 109(2), 415-438.

Campbell, Donald T. 1969. "Reforms as Experiments," *American Psychologist,* 24: 409–429.

Card, David, and Alan Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy*, 100 (1): 1-40.

Card, David, and Alan Krueger. 1996. "Labor Market Effects of School Quality: Theory and Evidence." NBER Working Paper No. 5450 February, 1996.

Csapó Benő, Molnár Gyöngyvér and Kinyó László. 2009. "A magyar oktatási rendszer szelektivitása a nemzetközi összehasonlító vizsgálatok eredményeinek tükrében." *Iskolakultúra*, 2009/3-4

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2010. "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." NBER Working Paper No. 16381

Coleman, J. S., Ernest Q. Campbel, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld and Robert R. York. 1966. *Equality of Educational Opportunity.* U. S. Department of Health, Education and Welfare.

Finn, Jeremy D., and Charles M. Achilles. 1990. "Answers and Questions about Class Size: A Statewide Experiment," *American Educational Research Journal,* 27, 557–577.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2013. Long-term effects of class size. *The Quarterly Journal of Economics,* 128(1), 249-285.

Hámori Bianka and Köllő János. 2011. "Kinek használ az évvesztés? Iskolakezdési kor és tanulói teljesítmények Magyarországon." *Közgazdasági szemle*, 58 (2), 133–157.

Hanushek, Eric. A. 1986. "The Economics of Schooling: Production and efficiency in Public Schools." *Journal of Economic Literature*, 24 (3), 1141 - 1177.

Hanushek, Eric A. 1997. "Assessing the effects of school resources on student performance: An Update." *Educational Evaluation and Policy Analysis*, 19(2), 141-64.

Heckman, James J.,A. Layne-Farrar, and Petra Todd. 1995 "Does Measured School Quality Really Matter?An Examination of the Earnings-Quality Relationship," NBER Working Paper No. 5274

Hoxby, Caroline M. 2000. The Effects of Class Size on Student Achievement: New Evidence from Population Variaton. *The Quarterly Journal of Economics,* 115 (4): 1239-1285.

Jepsen, Christopher, and Steven Rivkin, "Class Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size." *Journal of Human Resources*, 44 (1), 223-250.

Kertesi Gábor and Kézdi Gábor. 2005. "Általános iskolai szegregáció, I. rész." *Közgazdasági Szemle*, 52 (4), 317–355.

Kertesi Gábor and Kézdi Gábor. 2009. "Általános iskolai szegregáció Magyarországon az ezredforduló után." *Közgazdasági Szemle*, 56 (10), 959–1000.

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics,* 114 (2), 497–532.

Krueger, Alan B., and Diane Whitmore. 2001. "The effect of attending a small class in the early grades on college testtaking and middle school test results: Evidence from Project STAR." *Economic Journal*, 111, 1-28.

Lazear, Edward P. 1999. "Educational Production." NBER Working Paper No. 7349

Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. 1999. Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis*, 21(2), 165-77.

Mosteller, Frederick. 1995. “The Tennessee Study of Class Size in the Early School Grades,” *The Future of Children: Critical Issues for Children and Youths,* 5 (2): 113–127.

Oktatási Hivatal (Hungarian Bureau of Education). 2010. “Változások az országos kompetenciamérés skáláiban.” Accessed at http://www.oktatas.hu/pub_bin/dload/kozoktatas/meresek/orszmer2010/valt_orszmer_skala_110228.pdf. Last retrieval 2 June, 2015.

Oktatási Hivatal (Hungarian Bureau of Education) 2014. The list of closed and intermittent schools. Accessed at: http://www.oktatas.hu/kozneveles/kir. Last retrieval June 1, 2015.

Rivkin, Steven G., Eric A. Hanushek and John F. Kain. 2005. “Teachers, Schools and Academic Achievement.” *Econometrica*, 73 (2), 417–458.

Schanzenbach, Diane Whitmore. 2014. “Does Class Size Matter?” National Education Policy Center. Accessed at http://greatlakescenter.org/docs/Policy_Briefs/Schanzenbach_ClassSize.pdf

Urquiola, M. 2006. “Identifying class size effects in developing countries: Evidence from rural Bolivia.” *Review of Economics and Statistics*, 88(1), 171-177.

Urquiola, Miguel, and Eric Verhoogen. 2009. "Class-Size Caps, Sorting, and the Regression-Discontinuity Design." *American Economic Review*, 99(1): 179-215.

Wooldridge, Jeffrey M. 2002. *Introductory Econometrics: A Modern Approach*. South Western.