Quasi-Experimental Estimates of Class Size Effect in Primary Schools in Poland

Jakubowski, Maciej and Sakowski, Pawel

Wydzial Nauk Ekonomicznych Uniwersytet Warszawski

15 March 2006

Quasi-Experimental Estimates of Class Size Effect in Primary Schools in Poland[1]

Maciej Jakubowski
mjakubowski@uw.edu.pl

Paweł Sakowski
sakowski@uw.edu.pl

discussion paper
march 2006

Faculty of Economic Sciences
Warsaw University
ul. Długa 44/50, 00-241, Warsaw, Poland

Abstract

In this paper we analyze class size effects in the case of primary schools in Poland. We use two empirical strategies to avoid endogeneity bias. First, we use average class size in a grade as an instrumental variable for actual class size. This allows us to control for within school selection of pupils with different abilities to classes of different sizes. Additionally, we estimate fixed effects for schools to control for differences between them. Second, we exploit the fact that there is an informal maximum class size rule. We estimate class size effect only for those enrollment levels where some schools decide to add a new class and thus dramatically lower class sizes. For such enrollment levels variance of class size is mainly exogenous and we argue that this allows to estimate quasi-experimental class size effects. In this case we again use average class size as an instrument with enrollment as a key control variable. Using both strategies we obtain similar findings. We found that the positive effects observed with OLS regression disappear when we use instrumental variables. If we avoid endogeneity bias, then class size negatively affects student achievement. However, this effect is rather small. We discuss methodology, possible bias of results and the importance of our findings to current policy issues in Poland.

JEL: I2, H52, C31

Keywords: class size, educational achievement, student sorting, school fixed effects, instrumental variables, regression discontinuity design

[1] Final version of this paper was published in the International Journal of Educational Research, 2006, vol. 45 no. 3, p202-215.

1 Introduction

Class size reduction is one of the most heavily discussed and controversial issues in educational debates all over the world. It is a popular policy proposal probably because of its simplicity. The common intuition is that in smaller classes teachers can devote more time to each pupil and it is easier for them to maintain order. Thus, intuitively, reduction in the average class size should improve student achievement. This makes class size reduction attractive to parents and policy makers. Additionally, teachers may prefer smaller classes because of a lower workload and a higher demand for their work. In countries like the United States or France the average class size is now much smaller than 30 years ago. Some people argue that this expensive policy has no effect on student achievement, but others still regard class size reduction as cost-effective (see Hanushek, 2003; and Krueger, 2003).

However, in many countries class size reduction is not an issue. This is because of scarce resources or a greater focus on more politically relevant matters. In fact, one can easily imagine a situation where the average class size is growing because of budgetary cuts. It is also possible that governments prefer to invest in other school resources like teacher training, computers etc. Governments can decide to save funds for these investments by raising the pupils per teacher ratio. It is also possible that reforms of the educational systems can affect average class size without explicit intention to do so. The problem here is that when we do not know the effect of a change in class size on teaching quality, then the overall result of reform is hard to predict.

In Poland, class size reduction is not a widely recognized policy issue. Other problems and questions are more heavily discussed, especially the financial problems of local governments and their growing independence. The Polish educational system was decentralized in the 1990s and increasingly local governments more independently use their power to organize local school systems. Local authorities share the financial burden of the provision of primary and secondary education with the central government. In some places subsidies from the central budget are sufficient, but in others they are not. Many “gmina” (the lowest level of local government, which is responsible for primary education) spend as much as 50% of their budget on education and try to cut growing costs without any proper investigation or knowledge of how these actions influence the quality of teaching. In rural areas many local authorities close smaller schools to save money for other expenses.

In the beginning of the reform period in the 1990s, the Ministry of Education was supposed to establish so-called "educational standards" which were expected to regulate issues like maximum or minimum class size and the pupils per teachers ratio. However, such regulations have never been introduced. Additionally, in Poland there is no research that can be used as a reference point for decision makers. Discussion about educational standards is heavily influenced by the political battle between teachers' trade unions and representatives of local governments. Only a few people concentrate on the quality of education rather than on the financial controversy alone. The aim of this study is to analyze the effect of class size to support public discussion with some data. This seems to be important because in the near future average class size could grow because of the financial problems of local governments.

The causal effect of class size on pupil achievement is empirically difficult to measure. Actual class size considerably varies between regions, cities, rural and urban areas. Different schools with different characteristics have different class sizes. There are also visible differences in class sizes within schools. These differences are very often correlated with important characteristics of pupils, parents, teachers, schools and local communities that heavily influence student achievement. In the case of free choice, parents who care more about their children's education will send them to better schools. One can also expect better schools to have bigger classes because of higher demand. Different class sizes within schools are sometimes correlated with student ability. For example, school principals may decide to place insubordinate pupils in smaller classes or they may place pupils whose parents financially support schools in smaller classes with additional teaching equipment. Local governments can additionally support chosen schools or establish rules for maximum or minimum class sizes and so one can expect to find smaller classes in richer communities. All these examples suggest that the actual class size is a result of many independent and unobserved choices and that it is almost impossible to control for all variables that could be correlated with student achievement and class size (for a thorough discussion of potential bias see Hoxby, 2000).

In other words, the problem of endogeneity of class size casts doubt on using simple OLS regression to estimate class size effect. That is why some scholars dispute earlier research that claims to have shown no relationship between school resources and student achievement (see discussion between Hanushek, 2003 and Krueger, 2003). They believe that one needs sophisticated methodology or experimental data to find unbiased estimates of class size effects.

One way to avoid problems caused by class size endogeneity is to use a proper instrumental variable. In the case of class size many scholars used as an instrument the grade-average class size in a school. This allow them to focus on the class size variation that is not caused by within-school selection of students with different abilities among classes of different size (see Akerhielm, 1995). However, average class size is often correlated with other variables that could affect student achievement. For example, average class size can be higher in more popular schools or in cities (important in the Polish case). This makes estimates based on this instrument suspicious even when one uses a full set of control variables. Having this in mind, scholars use other techniques to allow for bias caused by between-school sorting.

One way to deal with this issue is to estimate fixed effects of schools or communities. This should give less biased estimates but one needs an appropriate data set (for example Hoxby (2000) uses longitudinal data on enrollment and achievement and Woessman (2003) uses scores from different grades). Another way to deal with a between-school bias is to find more credible sources of exogenous variation in the class size. Caroline Hoxby exploits random changes in cohort sizes to estimate class size effects (Hoxby, 2000). She also exploits the fact that random changes in enrollment cause dramatic change in a class size when a new class is added. Joshua Angrist and Victor Lavy use this idea to construct an instrument based on the maximum class size rule for the Israeli case and to obtain quasi-experimental estimates of class size effect (Angrist, Lavy, 1999).

We adopt similar ideas to examine class size effect in primary schools in Poland. We use 2002-2004 test scores from the obligatory primary school leavers' exam. We focus on two empirical strategies which we believe give unbiased results. The first strategy uses schools' grade-average class size as an instrument for an actual class size. It allows us to avoid bias caused by the within-school segregation of pupils with different abilities to classes of different sizes. As we have a three-year data set, we are able to control for schools' fixed effects. Basically, we look at how change in the average class size affects achievement in each school between years. In the second strategy, we use the fact that average class size changes abruptly when the higher enrollment level forces school principals and local authorities to increase the number of classes in a school. The idea is that the change in the number of classes depends mainly on the cohort size and is believed to be exogenous. Thus, we estimate class size effect only for those enrollment levels where the variance in class size can. to a large extent, be attributed to the change in the number of classes. We use two instruments in this case. One is identical to that developed by Angrist and Lavy. The second one is again the average class size. We discuss the difference between these two instruments in this case. We argue that exploring different methods and sources of exogenous variation give us a valuable cross-checking on empirical approaches.

The rest of the paper is organized as follows. In section 2 we depict the educational system and discuss former research on similar issues in Poland. In section 3 and 4 the data set and methodology are described. In section 5 results and simple application are presented. Section 6 summarizes our work.

2 Background of the study

There are two levels of compulsory education in Poland: 6-year primary school for 7-13 years of age and gymnasium (3-year lower secondary school) for 13-16 years of age. In this work we focus on public primary schools. These schools are free of charge and the only admission criterion is the age limit. Parents have a right to register their children in the nearest school but they can also try to find a place in another public or private institution. However, neither of these types of institutions are obliged to offer admission and private schools will usually charge tuition. For most Polish families private schools are too expensive and only a small proportion of pupils use them (less than 2% in 2004).

During the 1990s, the Polish educational system was decentralized. Local governments receive from the central government special grants calculated on a per pupil basis. These monies can be spent on schools, but also on other expenditures. Some local authorities spend additional money on schools from their own revenues while others are able to lower costs and to spend resources on other expenditures. However, 80% of school costs are attributed to teachers' salaries which are defined centrally. The easiest way to lower cost is to close smaller schools and to make bigger classes. Thus, local authorities have a great financial incentive to increase average class size and some of them do their best to do this.

Between 1990 and 2004, 5189 primary schools were closed. This is partly due to the decline in the overall number of pupils and the introduction of 6 years primary schools and 3 years gymnasiums (instead of 8 years primary) but also to the 'rationalization' of the school network, which was promoted by the financial incentives described above. The average class size in newly established gymnasiums is much higher than in primary schools (25 comparing to 21) and these schools are bigger (on average 242 pupils, compared to 186 pupils; data from 2004). In cities these numbers are higher and there are regions where the average class size is around 30. Some scholars believe that this is the proper way to save money for investment in more important resources (see Levitas, Golinowska, Herczynski, 2001). However, they do not analyze quantitatively how resources are related to teaching quality.

The fact is that while average class size in primary schools in Poland remains quite stable, during the last years the discrepancies between local governments have been substantial. While there are about 200 local governments out of 2622 with an average class size equal to or lower than 15, there is a similar amount with more than 30 students in class on average. Differences at the school level are even bigger.

Despite political discussion about teaching quality in Poland there are only a few papers that analyze the effects of school resources using quantitative methods. A few papers have touched the problem of class size effects, however none deal with the endogeneity issues. They use simple correlation or regression analysis to find positive or statistically insignificant relations between class size and achievement (see Bialecki, Haman, 2003; Herczynski, Herbst, 2002; Sleszynski, 2002). While positive effects can be explained by peer effects (see Dobbelsteen et al., 2002) we do not trust results from these studies because of their poor methodology.

In our earlier work we analyzed data sets with much more detailed information on schools than these that were available in the Polish research mentioned above (Jakubowski, Sakowski, 2004). To avoid endogeneity we also used average class size as an instrumental variable. We observed that the initial positive effect of class size became smaller when we used the instrumental variable. However, we were sure that in the Polish case this instrument alone is not sufficient because the average class size is also correlated with important characteristics of local governments, communities and schools. We were deeply unsatisfied with these results and believed that our empirical strategy and data does not allow us to estimate unbiased class size effect. This is the reason why we conducted the study described below.

3 Data

In our study we use data on student achievement on standardized exams conducted in Poland at the end of primary school. Starting from 2002, every pupil in the 6th grade of primary school has to solve a test designed to check his/her knowledge and skills needed to continue education at a higher level. The exam is obligatory and exceptions are very rare.

In this study we concentrate on the biggest region in Poland - Mazowieckie Voivodship - where about 13% of all Polish students live. This region is contains economically and socially differentiated areas. Warsaw, the capital of Poland, and its suburban area are the most prosperous regions of the country, while at the same time rural areas of Mazowieckie are among the poorest areas in the European Union. One has to keep in mind that these differences are probably correlated with class sizes in schools.

We received data on exam results for all primary schools in Mazowieckie from the Regional Examination Board in Warsaw. The data were aggregated on the class level with variables describing class mean score, the number of students in the class and indicators for schools. Enrollment was calculated as a sum of class sizes in each school. Because the exam is obligatory these numbers are very close to the real number of students in the class or school. At the time the research was carried out data from the years 2002 to 2004 were available.

In the original data set we had 9462 classes. We omitted privately run schools (304 classes), public schools for children with special needs, as well as schools with music and sport classes, those run in hospitals or other unusual places and finally schools for adults (in total 324 classes). These classes were irrelevant because our focus was on regular public schools. Additionally, from 8834 classes in public schools we excluded 143 classes with 40 or more students, because they cannot be run legally, so we treated them as an error. We also decided to eliminate schools which have classes with 10 or less pupils for enrollment levels greater than 20 (27 cases). Visual inspection of eliminated data shows that omitted schools were either incorrectly coded or have classes for pupils with special needs. Thus, our final data set contains 8664 classes (in total we eliminated 8,4% of the whole sample). However, we repeated all estimations on the full sample to find similar results. In Table 1 we show descriptive statistics of variables used in the analysis.

4 Methodology

Our methodology focuses on empirical strategies that allow us to avoid endogeneity bias. The dependent variable in each model is the class mean score in the primary school 6th graders exam. We start with a general model of the following form:

$$T_{kst} = \alpha + \beta_C c_{kst} + \mathbf{K}'_{kst}\,\boldsymbol{\beta}_{\mathbf{K}} + \mathbf{S}'_{st}\,\boldsymbol{\beta}_{\mathbf{S}} + \mathbf{Y}'_t\,\boldsymbol{\theta} + \varepsilon_{kst} \qquad (1)$$

where the dependent variable T_{kst} is a class average score, variable c_{kst} denotes the logarithm of actual class size, the column vectors $\mathbf{K}_{kst}$ and $\mathbf{S}_{st}$ contain, respectively, class-level (k) and school-level variables (s) and finally $\mathbf{Y}_t$ is a set of dummy variables for years of observation (t). We are interested in estimation of the coefficient β_C. Similar to Hoxby (2000) we measure class size in natural logs because a change of class size by one student is proportionally larger for smaller classes. Reducing a 15-student class by 5 students could have a stronger effect than reducing a 30-student class by the same amount.

As we said earlier such specification of the education production function and its estimation using the *ordinary least-squares* method results in biased estimates of β_C. This is because class size is endogenous, which means that it is correlated with many unmeasured, unobserved or simply omitted factors. Possible reasons for such endogeneity are the between-schools and within-school sorting of students into classes of different sizes. Including large set of control variables describing students, parents and community characteristics in the equation (1) is not a satisfactory solution. One also needs to incorporate a complete set of lagged inputs - for example class sizes that students of a particular class experienced or resources they experienced in their pre-school education. In practice this is almost impossible and one will always come across some sort of bias using OLS (see Todd and Wolpin, 2003). The common problem in the production function approach is that students with different abilities, which are very difficult to measure, cluster in classes of different sizes.

One of the methods to avoid endogeneity bias is to apply the *instrumental variable* technique using the *two-stage least-squares* method. However, this is restricted by the narrow set of possible instruments[2]. In this paper we discuss two instruments already used in several studies.

Our first approach is based on using as an instrument for c_{kst} the natural logarithm of the schools' grade-average class size $\bar{c}_{kst}$ (Akerhielm, 1995; Wößmann, 2003). This way we avoid bias from the selection of students of different abilities to classes of different sizes. As far as there is no segregation between schools or one fully controls for it, the average class size is believed to be exogenous and can be used to obtain unbiased estimates.

The first step to implement 2SLS method is to estimate the equation:

$$c_{kst} = \varpi + \beta_{\bar{c}}\bar{c}_{kst} + \mathbf{K}'_{kst}\,\boldsymbol{\beta}_{\mathbf{K}} + \mathbf{S}'_{st}\,\boldsymbol{\beta}_{\mathbf{S}} + \mathbf{Y}'_{t}\,\boldsymbol{\theta} + \eta_{ksgt} \tag{2}$$

where the class size logarithm c_{kst} is explained by the logarithm of the average class size in a grade $\bar{c}_{kst}$ and all other exogenous variables. Secondly, having substituted theoretical values $\hat{c}_{kst} = c_{kst} - \eta_{kst}$ from the first equation we estimate the following regression:

$$T_{kst} = \alpha + \beta_{\hat{c}}\hat{c}_{kst} + \mathbf{K}'_{kst}\,\boldsymbol{\beta}_{\mathbf{K}} + \mathbf{S}'_{st}\,\boldsymbol{\beta}_{\mathbf{S}} + \mathbf{Y}'_{t}\,\boldsymbol{\theta} + \varepsilon_{kst} \tag{3}$$

where $\beta_{\hat{c}}$ is believed to be an unbiased estimate of the class size.

However, as we said already this approach results in unbiased estimates only when one fully controls for differences between schools and there are several reasons for between-school sorting to be strongly correlated with class size. We mentioned some reasons earlier, but just to reiterate a few of them: parents' decisions concerning residential place are linked to their children's abilities and school resources. With school choice, better schools will have higher enrollment levels etc. In Poland schools in rural areas are much smaller and at the same time they scored lower on external exams, which is a very important source of bias. One way to avoid such bias is to estimate the *fixed effects* of schools.

2 An instrument has to be chosen very carefully. Using wrong instruments can result in even more biased estimates (Angrist, Krueger, 2001).

To apply this idea we include in the model equation the instrumented class size and complete set of school dummies $\mathbf{S}_s \boldsymbol{\varphi}$:

$$T_{kst} = \alpha + \beta_{\hat{c}} \hat{c}_{kst} + \mathbf{S}'_s \boldsymbol{\varphi} + \mathbf{Y}'_t \boldsymbol{\theta} + \varepsilon_{kst} \qquad (4)$$

Use of the 2SLS method, with the average class size and school dummies, gives us the opportunity to focus on class variation that is caused by fluctuations between years in enrollment levels in the same school. In this approach we follow research by Wößmann and West, who claim this source of variation to be exogenous (Wößmann and West, 2002; Wößmann, 2003). However, as Hoxby points out (2000) in this method one has to exclude class size changes that are unusual, too big or too small, because they could be caused by external factors (e.g. high migration). Thus, we also estimate school-fixed effects 2SLS in a sub-sample with limited magnitude of class size changes.

Our second approach exploits the idea that there is a dramatic change in a given school's average class size when a new class is added. Such variation in class size is believed to be exogenously caused if there exists a rule that defines maximum class size in public schools. One way to apply this idea is to calculate the function:

$$C^*_{st} = f(E_{st}) = \frac{E_{st}}{\mathrm{int}(\frac{E_{st} - 1}{C^{\max}}) + 1} \qquad (5)$$

and to use a logarithm of this function as an instrument of the class size C_{kst}

$$c^*_{st} = \ln(C^*_{st}) \qquad (6)$$

where C^*_{st} denotes theoretical class size given enrollment in a particular school E_{st} and $C^{\max}$ is the maximum class size. Following Angrist and Lavy, we call the equation (6) a *theoretical class size function* (Angrist, Lavy, 1999). Using this function as an instrumental variable gives us a setting similar to the regression-discontinuity design in its fuzzy version since some schools add new classes earlier and some do it later (Lee, 2005).

This instrument is based on the assumption that a formal or commonly obeyed rule that defines maximum class size exists. Angrist and Lavy used a so called "Maimonides Rule" that in Israel limits the maximum number of students in a class to 40. As we mentioned earlier, there is no formal regulation of class size in Poland. However, there is a strong relationship between grade enrollment and class size. In Figure 1 one can easily notice enrollment levels for which in some schools new classes are being created. It seems that most schools in Poland add new classes when the average class size is about 28 or 29. Except threshold levels, this graph is very similar to that presented for Israel in Angrist and Lavy paper. Similarly, only in the first two or three discontinuity samples is the relationship between actual class size and theoretical class size function (with 29 as the maximum class size in this example) strong and visible. This fact, and the relatively high correlation coefficients in the whole sample, let us believe that the theoretical class size determined by function (6) could be a valid instrument for actual class size in our case[3].

It has to be noticed that such theoretical class size function is identical to the average class size when the number of classes in each school is similar. Due to this, one needs to fully control for differences between schools with different enrollment levels. Where this is not the case, this instrument provides us with unbiased quasi-experimental estimates only in the neighborhood of the enrollment threshold level where new classes are added[4].

It is worth mentioning here the difference between the two instruments described above. When the number of classes in schools is the same then these instruments are identical. In other case, where there is a discontinuity between enrollment and class sizes, they work very differently. When one uses grade-average class size in the discontinuity sample then one has to control for the differences between schools with different numbers of classes or to assume that they are negligible. The wider the discontinuity sample, the more critical are such differences. Schools that 'too early' create a new class are probably different than schools that do not. They could be placed in richer or more influential communities or simply have more motivated parents and principals.

The crucial point here is the choice of discontinuity sample. If it is chosen properly then schools with unusually large and small class sizes are excluded. So, one can assume that

3 Theoretical class size functions constructed for maximum class sizes of 28 and 29 explain about 60% of class size variation in our data set. This number is very similar to that obtained for Israel by Angrist and Lavy (1999).
4 The point is that one needs a really large sample to obtain precise estimates when one wishes to conduct an analysis only for such "discontinuity samples" (see Hoxby, 2000).

differences in class sizes are mostly due to bureaucratic decisions and not to parental or school characteristics. Where this is the case, using theoretical class size function with arbitrary chosen maximum class sizes seems suspicious, because we treat similar schools differently. For example, we have three public schools with similar enrollments near the threshold level. In Poland these could be schools with enrollments like 56, 57 and 58 students. Now, assume that the smallest school creates a new class and two others do not. The reason could be that in the smallest school there were three students who initially enrolled but then decided to move to another place. Applying theoretical class size function in this case seems implausible.

Instead one can use average class size as an instrument with enrollment as a key control variable. We believe that this is a proper instrument in discontinuity sample where the variance in class sizes is mainly exogenous. The benefit derived from such an approach is that in our case the average class size is more closely correlated with actual class size than theoretical class size. The cost is that schools which add new classes for lower enrollment levels can be different than schools that add new classes for higher enrollment levels and in our case we do not have any additional data that can be used to check this. Instead we use enrollment as a key control variable and test robustness of our results for different discontinuity samples.

5 Results

We start with results calculated separately for each year (Table 2). OLS estimates without any control variables show that class size is positively correlated with student achievement. Adding enrollment considerably lowers class size coefficients, although, they still remain positive. Note that in all regressions enrollment is positively correlated with achievement. Results obtained for each year with 2SLS are quite similar. Clearly, estimated coefficients of class size remain positive or insignificant.

To obtain unbiased estimates we conduct regressions with data pooled over three years. Results are presented in Table 3[5]. The first column shows that OLS and IV methods indicate positive effects, which repeats observations from one-year samples. However, in this case we were able to estimate school fixed-effects to focus on the effect of changes in average class size between years in one school. As we said earlier, such changes should give a credible source of exogenous variation. These crucial results are shown in the second column of Table 3. We see that in both cases IV method with school-fixed effects evidences a negative relationship between class size and achievement. Additionally we repeated analysis for schools for which a change in the number of classes was 0 or +/-1. This was made to assure that we look at a class size variation that is really exogenous[6]. Results in column (4) show that this way we obtained similar results.

Our second empirical strategy was to use IV method only for discontinuity samples. In Table 4 we present results obtained for 29 as the maximum class size and +/-5 neighborhood (columns 1 and 2) and +/-4 neighborhood (columns 3 and 4). We also repeated our investigations for different samples to obtain similar results[7]. We analyzed only the first three discontinuity samples, for thresholds of 29, 58 and 87, because for higher enrollment levels the patterns of class size change were less clear.

Table 4 shows that in the case of the OLS method almost all class size coefficients are insignificant. Crucial results are shown in the 2nd and 4th columns where results from IV methods with control for enrollment are presented. Thus, we concentrate on differences between schools with the same enrollment levels but different numbers of classes. We see that,

[5] In all regression with pooled data we introduced dummies for years. We did not report them, however, there were significant and correctly reflect year after year decline in average scores.

[6] We also tried with limits set to 20 and 30% change of enrollment to find similar results.

[7] Additional results can be found in the working paper available at www.wne.uw.edu.pl/mjakubowski or can be

regardless of the method, without controlling for enrollment all coefficients are positive. When we control for enrollment all IV coefficients become negative, however, in the case of theoretical class size function they are insignificant. One has to notice that in the latter case standard errors are considerably bigger than in earlier regressions. Note also that in all cases enrollment is positively correlated with achievement.

The results presented above indicate negative class size effects when we use credible sources for variation. They indicate that a class size reduction by 10% improves class mean score by no more than 0.08 points. This is less than 5% of the standard deviation of class mean scores. Compared to the results of other experimental and quasi-experimental studies, this number is smaller (see Angrist, Lavy, 1999; Krueger, 2003; Urquiola, 2000)[8]. One can say that this effect is negligible from the practical point of view. Therefore, we applied our results to one of the most controversial issues in educational policy in Poland, namely, to the widening gap between student achievement in urban and rural areas.

As we said at the beginning of this paper, some scholars believe that rural local governments should increase class sizes to save money for investments in other resources (see Levitas, Golinowska, Herczynski, 2000). No one shows quantitatively what will be the effect of these "other" investments, however, we can apply our results to calculate how raising average class size in rural areas to a level similar to that found in urban areas can influence student achievement. According to our estimates implementing such a policy would widen the gap between exam scores in urban and rural areas by 13 to 47 percent depending on the year and which estimate we used[9]. This number is not negligible. Calculations show that even when effects obtained in this research seem to be small when measured by exam points, they are important for policy concerns.

---

requested by email: mjakubowski@uw.edu.pl.

[8] Hoxby (2000), using similar methods, shows that there is no class size effect. However, she uses data from Connecticut, USA, where class sizes are considerably smaller.

9 We calculated how increasing class sizes in rural schools will affect the difference in achievement between rural schools and schools in cities. We exluded Warsaw, which each year scores 2-3 points higher than the rest of the country and has exceptional economic and social characteristics. Calculations were based on class size means for each year separately. In almost all cases differences and changes in differences were statistically significant (the only exception is the lower estimate for year 2004). One can easily repeat these calculations using data presented in the lower part of Table I.

6 Conclusion

This paper analyzed class size effects in the case of primary schools in Poland. In addition to OLS regressions, 2SLS method with instrumental variables was applied. Two different instruments were used: the school grade-average class size and the theoretical class size function based on a maximum class size rule. The differences between these instruments were discussed.

Two empirical strategies were applied. The first strategy was to estimate class size effects using 2SLS method with instrumental variables and school fixed effects. Results indicate that a between-years variation in class sizes in schools negatively affects student achievement. The claim was made that such variation is mainly exogenous. The second strategy focused on an analysis of discontinuity samples. Again, 2SLS method with instrumental variables was applied. Obtained estimates again indicated that class size negatively affects student achievement.

The obtained results are of a practical importance. In recent years average class size has not changed in Poland. However, many local governments try to increase class sizes, especially in rural areas where they experience financial problems caused by educational expenses. Thus, in many rural schools class size might grow considerably in the near future. This paper has shown that increasing class size in rural schools to the level of urban schools might further widen the achievement gap between them by 13 to 47 percent. These results show that class size effects should be widely recognized by scholars and policy makers, because they significantly influence achievement.

References

Akerhielm, K. (1995). Does Class Size Matter? *Economics of Education Review, XIV,* 229-241.

Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: from Supply and Demand to Natural Experiments. *Princeton University, Working Paper 455*.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics*, 533-575.

Bialecki, I., & Haman, J. (2003). *Program Miedzynarodowej Oceny Umiejetnosci Uczniow OECD/PISA. Wyniki polskie - raport z badan.* IFiS PAN, Warsaw.

Hanushek, E. A. (2003). The Failure of Input-based Schooling Policies. *Economic Journal 113 (485):* F64-F98.

Herczynski, J., & Herbst, M. (2002). *Pierwsza odslona.* Fundacja Klub Obywatelski.

Hoxby, C. M. (2000). The effects of class size and composition on student achievement: new evidence from natural population variation. *Quarterly Journal of Economics, 115(4)*, 1239-1285.

Jakubowski, M., & Sakowski, P. (2004). *Wpływ wielkości klasy na wyniki uczniów w szkołach podstawowych.* Mimeo, Department of Economics, Warsaw University.

Kruger, A. B. (2003). Economic Considerations and Class Size. *Economic Journal 113 (485)*: F34-F63.

Lee, M. (2005). *Micro-Econometrics for Policy, Program and Treatment Effects,* Series: Advanced Texts in Econometrics, Oxford University Press

Levitas, A., & Golinowska, S., & Herczynski J. (2001). *Improving Rural Education in Poland*. CASE - Center for Economic and Social Studies, Warsaw.

Sleszynski, P. (2002). *Ekonomiczne uwarunkowania wynikow sprawdzianu szostoklasistow i egzaminu gimnazjalnego przeprowadzonego wiosna 2002 roku.* Ministry of Education, mimeo.

Todd, P., & Wolpin, K. (2003). On the Specification and Estimation of the Production Function of Cognitive Achievement. *Economic Journal 113 (485).*

Urquiola, M. (2000). *Identifying class size effects in developing countries: Evidence from rural schools in Bolivia*. Development Research Group The World Bank.

Wößmann,, L. 2003. European Education Production Functions: What Makes a Difference for Student Achievement in Europe? *European Economy, Economic Papers* No. 190, European Commission, Brussels.

Wößmann, L., & West, M. R. (2002). Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS. *Harvard University, Program on Education Policy and Governance Research Paper,* PEPG/02-02.

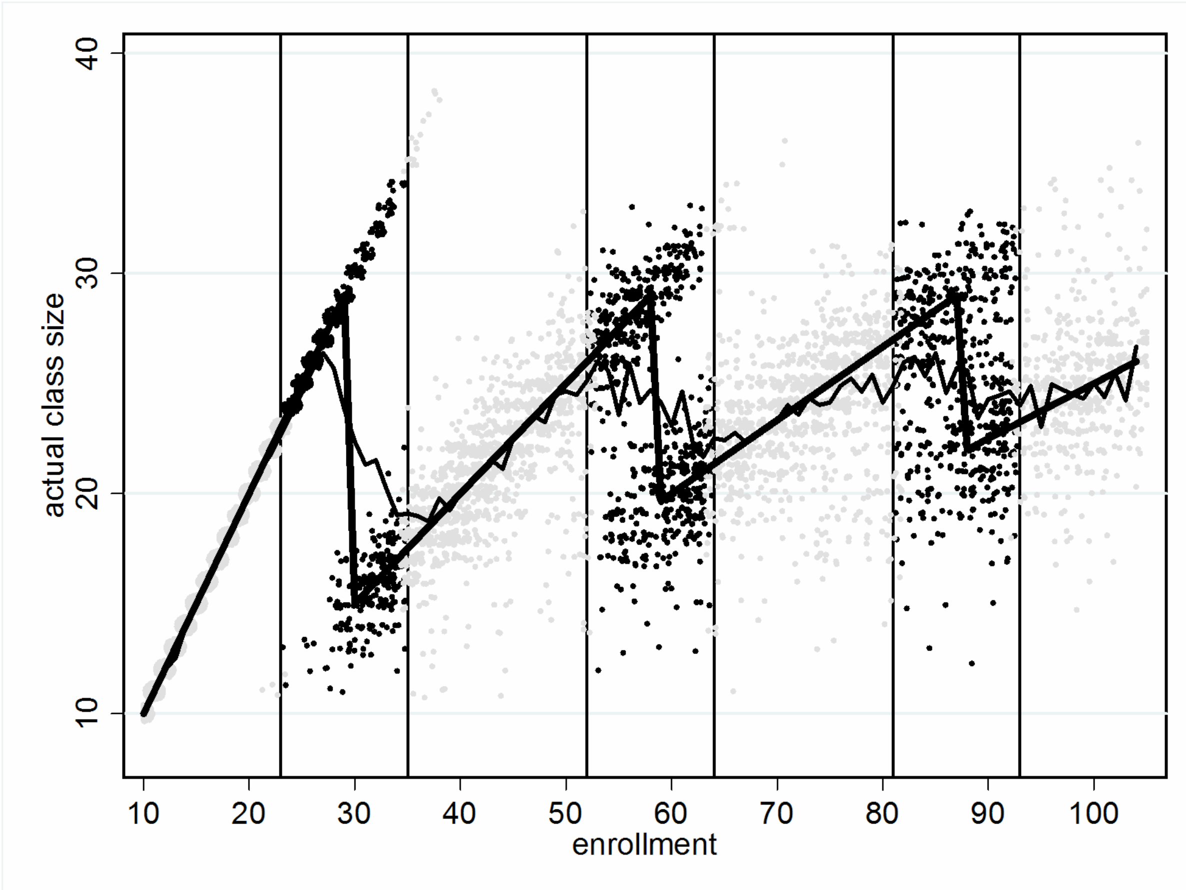

Figure 1. Actual class size, theoretical class size function, grade-average class size and discontinuity samples. Pooled sample, 2002-2004.

Thick solid line represents theoretical class size function (with threshold set to 29). Thin solid line stands for grade-average class size for given enrollment level. Each dot represents one class. In order to avoid dots overlapping random noise was added to the data. Discontinuity samples were constructed for following grade enrollment levels: 24-34, 53-63, 82-92.

Table 1

DESCRIPTIVE STATISTICS

	pooled sample, 2002-2004	discontinuity sample (1) 29+/-5*	discontinuity sample (2) 29+/-4*
Class size	21.05	24.03	24.18
	(6.55)	(5.12)	(5.23)
Enrollment	63.49	56.59	50.38
	(49.17)	(23.47)	(23.27)
Number of classes	8664	1867	1511
Number of schools	1657	658	561

	2002 sample			2003 sample			2004 sample		
	whole sample	urban without Warsaw	rural only	whole sample	urban without Warsaw	rural only	whole sample	urban without Warsaw	rural only
Class size	21.02	24.95	17.77	21.15	25.17	17.93	20.98	23.96	18.01
	(6.59)	(3.93)	(6.91)	(6.65)	(4.02)	(6.99)	(6.38)	(5.15)	(6.73)
Mean Score	30.36	30.41	29.76	27.56	28.13	26.48	25.92	26.15	24.68
	(2.89)	(2.83)	(2.88)	(3.00)	(2.55)	(2.87)	(3.43)	(3.06)	(3.09)
Number of classes	2891	651	1206	2903	648	1218	2852	700	1124

Standard deviations are reported in parentheses.
* Discontinuity samples were constructed for following grade enrollment levels:
(1) 24-34, 53-63, 82-92, (2) 25-33, 54-62, 83-91

Table 2

ESTIMATION RESULTS

YEARLY SAMPLES (2002,2003,2004)

	2002		2003		2004	
Regressors: ↓	(1)	(2)	(3)	(4)	(5)	(6)
OLS ESTIMATES						
Log. of class size	.475***	.164	1.315***	.515***	1.560***	.478**
	(.131)	(.159)	(.125)	(.147)	(.157)	(.188)
Enrollment	—	.004***	—	.012***	—	.016***
		(.001)		(.001)		(.001)
R^2	0.0041	0.0078	0.0364	0.0670	0.0328	0.0660
2SLS ESTIMATES — INSTRUMENTAL VARIABLE: GRADE-AVERAGE CLASS SIZE						
Log. of class size	.218	-.246	1.167***	.2348	1.396***	.145
	(.134)	(.164)	(.127)	(.152)	(.161)	(.194)
Enrollment	—	.006***	—	.014***	—	.017***
		(.001)		(.001)		(.001)
R^2	0.0028	0.0055	0.0359	0.0658	0.0324	0.0650
2SLS ESTIMATES — INSTRUMENTAL VARIABLE: THEORETICAL FUNCTION OF CLASS SIZE						
Log. of class size	.212	-.354	1.347***	.234	1.724***	.349*
	(.142)	(.181)	(.140)	(.176)	(.173)	(.217)
Enrollment	—	.006***	—	.014***	—	.017***
		(.0013)		(.001)		(.001)
R^2	0.0028	0.0042	0.0364	0.0658	0.0324	0.0658
N	2891		2903		2852	

Robust standard errors are reported in parentheses. One, two, and three asterisks denote significance at the 0.10, 0.05, and 0.01 levels, respectively. The unit of observation is the average score in the class.

Table 3

ESTIMATION RESULTS

POOLED SAMPLE (2002-2004)

	Full sample		Restricted sample *	
Regressors: ↓	(1)	(2)	(3)	(4)
OLS ESTIMATES				
Log. of class size	1.119***	.919***	1.113***	.797***
	(.079)	(.150)	(0.125)	(.247)
School dummies	NO	YES	NO	YES
R^2	0.2736	0.5628	0.2703	0.6522
2SLS ESTIMATES — INSTRUMENTAL VARIABLE: GRADE-AVERAGE CLASS SIZE				
Log. of class size	.930***	-.583***	0.941***	-.644***
	(.081)	(.173)	(.127)	(.234)
School dummies	NO	YES	NO	YES
R^2	0.2732	0.5566	0.2699	0.6477
2SLS ESTIMATES — INSTRUMENTAL VARIABLE: THEORETICAL FUNCTION OF CLASS SIZE				
Log. of class size	1.086***	-.755***	1.104***	-.746***
	(.087)	(.227)	(0.141)	(.297)
School dummies	NO	YES	NO	YES
R^2	0.2736	0.5551	0.2703	0.6470
N	8664		8414	

Robust standard errors are reported in parentheses. One, two, and three asterisks denote significance at the 0.10, 0.05, and 0.01 levels, respectively. The unit of observation is the average score in the class.

* In the restricted sample schools were excluded where the number of classes changed by more than one (from year to year).

Table 4

ESTIMATION RESULTS

DISCONTINUITY SAMPLES

	29 +/-5 24-34, 53-63, 82-92		29 +/-4 25-33, 54-62, 83-91	
Regressors: ↓	(1)	(2)	(3)	(4)
OLS ESTIMATES				
Log. of class size	.478 (.304)	.013 (.299)	.346 (.326)	-.0204 (.320)
Enrollment	—	.0302*** (.003)	—	.0289*** (.003)
R^2	0.2669	0.3053	0.2973	0.3317
2SLS ESTIMATES, INSTRUMENTAL VARIABLE: THEORETICAL FUNCTION OF CLASS SIZE				
Log. of class size	2.107** (.978)	-.469 (1.044)	2.251* (1.202)	-.242 (1.268)
Enrollment	—	.031*** (.003)	—	.0292*** (.004)
R^2	0.2557	0.3044	0.2815	0.3315
2SLS ESTIMATES, INSTRUMENTAL VARIABLE: GRADE-AVERAGE CLASS SIZE				
Log. of class size	-.073 (.321)	-.632** (.317)	-.139 (.344)	-.576* (.338)
Enrollment	—	.031*** (.003)	—	.029*** (.003)
R^2	0.2656	0.3036	0.2963	0.3304
N	1867		1511	

Robust standard errors are reported in parentheses. One, two, and three asterisks denote significance at the 0.10, 0.05, and 0.01 levels, respectively. The unit of observation is the average score in the class.