I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11208

Does Class Size Matter for School Tracking Outcomes after Elementary School? Quasi-Experimental Evidence Using Administrative Panel Data from Germany

Bethlehem A. Argaw
Patrick A. Puhani

DECEMBER 2017

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11208

Does Class Size Matter for School Tracking Outcomes after Elementary School? Quasi-Experimental Evidence Using Administrative Panel Data from Germany

Bethlehem A. Argaw
Leibniz Universität Hannover

Patrick A. Puhani
Leibniz Universität Hannover, CReAM, University College London, SEW, University of St. Gallen, and IZA

DECEMBER 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Does Class Size Matter for School Tracking Outcomes after Elementary School? Quasi-Experimental Evidence Using Administrative Panel Data from Germany

We use administrative panel data on about a quarter of a million students in the German state of Hesse to estimate the causal effect of class size on school tracking outcomes after elementary school. Our identification strategy relies on the quasi-random assignment of students to different class sizes based on maximum class size rules. In Germany, students are tracked into more or less academic middle school types at about age ten based, to a large extent, on academic achievement in elementary school. We mostly find no or small effects of class size in elementary school on receiving a recommendation or on the actual choice to attend the more academic middle school type. For male students, we find that an increase in class size by 10 students would reduce their chance of attending the higher school track – which more than 40 percent of students attend – by 3 percentage points.

JEL Classification: I21, I28
Keywords: class size, panel, administrative data, education production

Corresponding author:
Patrick A. Puhani
Leibniz Universität Hannover
Institut für Arbeitsökonomik
Königsworther Platz 1
30167 Hannover
Germany
E-mail: puhani@aoek.uni-hannover.de

1 Introduction

The policy of class size reductions has been at the center of the educational research and policy debate for a few decades. The impact of class size on educational achievements is found to vary, among other dimensions, across school systems, grade levels, gender and students' socio-economic background. Whereas some empirical studies find a positive effect of smaller classes on short- and long-term outcomes (Card and Krueger, 1996; Angrist and Lavy, 1999; Krueger, 2003; Dustmann *et al.*, 2003), part of the literature finds no substantial benefits from class size reductions (Hoxby, 2000; Levin, 2001; Dobbelsteen *et al.*, 2002; Hanushek, 2003; Wößmann, 2005). Estimation bias arising from non-random sorting of students into classes of different sizes within and across schools has been well studied using randomized experiments (Krueger, 1999; Krueger and Whitmore, 2001) or quasi-randomized experiments (Angrist and Lavy, 1999; Hoxby, 2000). Recently available administrative datasets, mainly from Scandinavian countries, revived the class size debate and improved the precision of previous results (Browning and Heinesen, 2007; Leuven *et al.*, 2008; Fredriksson *et al.*, 2013).

In this paper, we provide evidence on the causal effect of class size using administrative data from the German state of Hesse, where the maximum class size is 25 students, which is lower than in Sweden (30) or Israel (40) as studied in Angrist and Lavy (1999) and Fredriksson *et al.* (2013), respectively. The benefit of class size reductions in the context of an early school tracking system such as in Germany is scarcely studied. Although there exists some evidence on the relationship between class size and educational outcomes in Germany, to the best of our knowledge, Wößmann (2005) is the only study that uses a credible identification strategy to estimate the causal effect of class size in Germany. The author uses data from TIMSS (Trends in International Mathematics and Science Study) from 15 Western European countries and finds no substantial benefit of smaller classes in lower middle schools in most countries including Germany. Our paper complements Wößmann (2005) by using administrative data and focussing on elementary instead of middle schools. Previous evidence from other countries shows that smaller classes are most beneficial during elementary or pre-elementary education (Angrist and Lavy, 1999; Hoxby, 2000; Ding and Lehrer, 2010). Although the currently available evidence shows that smaller classes in German lower middle schools do not improve achievements, it still remains an open question whether or not class sizes in German elementary schools influence educational outcomes.

Much of the debate on the effectiveness of class size reduction to improve students' educational outcomes revolves around getting an unbiased and precise estimate of the benefit of

smaller classes. The methodological challenge in establishing a cause-and-effect relationship between class size and educational outcomes arises due to the non-random sorting of students between and within schools. To the extent that school and family background characteristics relevant for students' academic performance remain unobservable and are correlated with class size, ordinary least squares regression (OLS) gives a biased estimate of the class size effect. Ideally, the causal effect of class size could be identified using randomized experiments: in Tennessee's Project STAR (Student-Teacher Achievement Ratio) experiment, students who were randomly assigned to smaller classes from kindergarten through 3^{rd} grade made significant improvement in mathematics and reading test scores (Krueger, 1999) as well as in long-term outcomes such as the probability to take the ACT or SAT college entrance examinations, especially for minorities (Krueger and Whitmore, 2001).

In the absence of experimental data, studies have relied on quasi-experimental designs based on observational data. One of the most widely used quasi-experimental approaches in the class size literature, which we also follow in this paper, is pioneered by Angrist and Lavy (1999) and uses the exogenous source of variation in class size induced by Maimonides' rule of 40 students in Israel's schools. Two otherwise similar school entry cohorts are assigned into classes of different size as a result of the variation in total enrollment and the maximum number of students allowed in a class. In a fuzzy regression discontinuity framework, Angrist and Lavy (1999) use the rule-induced class size as an instrumental variable for actual class size. They find significant improvement in reading and mathematics test scores for students taught in smaller classes in 4^{th}h and 5^{th} grade. In another quasi-experimental approach, Hoxby (2000) uses the variation in class size across school entry cohorts arising from natural fluctuation in population size. The randomness in the timing of birth coupled with school entry age rules results in adjacent school entry cohorts starting school in classes of different size. Based on long panel data from Connecticut, Hoxby (2000) does not find any significant positive effect of small classes on math, reading and writing scores in 4^{th} and 6^{th} grade.

Quasi-experimental designs, especially based on maximum class size rules, have been applied to estimate the class size effect in different countries, at different school levels and using survey and/or administrative data. For instance, Browning and Heinesen (2007) for Denmark; Piketty (2004), Gary-Bobo and Mahjoub (2006) for France; Levin (2001) and Dobbelsteen *et al.* (2002) for the Netherlands; Bonesroenning (2003) and Leuven *et al.* (2008) for Norway; Fredriksson *et al.* (2013) for Sweden; Urquiola (2006) for Bolivia, Wößmann (2005), Wößmann and West (2006) for several Western countries. The evidence shows mixed results. Small class size is causally linked to higher test scores in Bolivia, France and Sweden, lower grade repetition in

France, higher years of education and wages in Sweden and Denmark. On the contrary, no significant improvement in test scores due to smaller classes is found in the Netherlands and Germany, whereas the evidence is mixed for Norway and the US.

Administrative data sets are becoming accessible mainly in Scandinavian countries and are being widely used for empirical research in recent years. For instance, Leuven *et al.* (2008) uses Norwegian administrative data that contains nationwide test scores and finds that the effect of class size in elementary school is almost zero. Using administrative panel data from Denmark, Browning and Heinesen (2007) find a small negative effect of larger classes in 8^{th} grade on the probability of completing secondary education. On the contrary, Fredriksson *et al.* (2013) uses Swedish administrative data matched with self-reported data on cognitive and non-cognitive skills and finds that individuals exposed to smaller classes in elementary school have substantially higher cognitive and non-cognitive skills, years of completed education and earnings as adults.

This paper contributes to the recent literature that relies on administrative data to identify the causal effects of class sizes. It does so by using school administrative data from the German state of Hesse to complement the existing evidence from Germany by Wößmann (2005). We measure academic performance based on the observation that students in Germany are tracked into more or less academic middle school types at the end of elementary school. These school types also referred to as school tracks, but in Germany students are tracked by segregating them into different schools, as described in Dustmann *et al.* (2016). The type of middle schools that students are recommended to attend and the school type they eventually attend, to a large extent, depends on their academic performance in Math, German and General Studies in elementary school. We use an indicator for getting a recommendation to attend the higher and more academic school type called *Gymnasium* and the actual choice to attend this type of middle school as the main measures of students' educational outcome. The panel nature of our data also allows us to consider the effect of class size in grade 1 or average class size in grades 1 to 4 of elementary school as impact variables as well as to define an indicator of grade repetition in elementary school as dependent variable.

The instrumental variable estimation results show either no or only small effects of class size in elementary school on receiving a recommendation or on the actual choice to attend the more academic middle school type. For male students, we find that an increase in class size by 10 students would reduce their chance of attending the higher school track (which more than 40 percent of students attend) by 3 percentage points. We also find that 10 more students in the 1^{st} grade of elementary school generally increase the chance of repeating a grade in elementary school by 4 percentage points, yet this results is only based on two school cohorts and should

be checked for robustness once a longer panel data set is available.

The rest of this paper is structured as follows. In Section 2, we describe the German school tracking system and the administrative teacher and student panel data. Section 3 discusses the empirical strategy and presents the main estimation results followed by sub-group analysis and robustness checks. Concluding remarks are provided in Section 4.

2 Administrative Data and School Tracking in Germany

We use the administrative teacher and student panel data (in German: *Lehrer- und Schüler-datenbank, LUSD*) that covers all students in the German state of Hesse for the school years 2007/08 until 2012/13.[1] It contains various measures of student-, teacher-, subject-, classroom- and school-level characteristics and has a panel nature. In this paper, we use the information on student characteristics such as age, gender, nationality, the grade and type of middle school recommendation and the actual type of middle school attended. In our main analyses, we restrict the sample to students who are in 4^{th} grade in each school year and, thanks to the panel component, we follow them into 5^{th} grade when they enter different types of middle schools. The data in the LUSD are collected in October or November, that is toward the beginning of each school year. The administrative panel data contains six school years. However, we lose observations for one school year since the outcome variable in school year t is measured based on data on middle school type in the following school year. Special education schools (*Sonderschule*, *Förderschule*) are excluded from the sample, but private schools are included.

We measure academic performance based on the observation that students in Germany are tracked into more or less academic middle school types at the end of elementary school. Unfortunately, the LUSD does not contain data on academic test scores. The first outcome variable takes the value one if a 4^{th} grader in year t gets a recommendation to attend the more academic school type (*Gymnasium*) and it takes the value zero if a student gets a recommendation to attend any of the non-academic school types. The second educational measure is defined in a similar notion but based on the actual middle school type that a student attends in 5^{th} grade. Whereas the recommendation is by statute strongly correlated with academic achievements in elementary school, the actual school type attended also captures the influence of parents in the tracking decision. Our full estimation sample contains 258,098 students during the five school years.

[1] The administrative data also covers previous school years 2002/03 to 2006/07. However, a time-consistent student identifier, which allows following students as they enter middle schools, was introduced in the 2007/08 school year.

The education system in Germany is decentralized across the 16 federal states (*Länder*). In most federal states, including the state of Hesse, students are tracked into more or less academic middle school types at about age ten, i.e., after four years in elementary school.[2] The most academic school type, called *Gymnasium*, lasts about nine years and it is the only school type that provides direct access to tertiary academic education (university or university of applied sciences). The intermediate (*Realschule*) and lower (*Hauptschule*) school types take six and five years respectively and provide qualification for entry into "dual education" where vocational schools are combined with apprenticeship training. Some share of students avoids tracking and enters comprehensive schools (*Gesamtschule*). In the German state of Hesse, there is also a possibility to postpone tracking until grade seven by attending the support stage (*Förderstufe*).

The type of middle school that students attend is determined based on their academic achievement in elementary school as well as on a judgment of their academic prospects in middle school. At the end of elementary school, students receive a school recommendation from primary school teachers, mainly based on their academic achievement in Mathematics, German and General Studies. In the state of Hesse, the track recommendation issued by the school and the parents' final track choice are determined by an iterative process involving consultative talks with the parents during the months February to April in the 4^{th} and final grade of elementary school. As in most German states, the track recommendation in the state of Hesse is not binding and hence parents make the final tracking decision. As a rule, it is possible to modify the initial school track choice later on in middle school. However, this is very uncommon mainly due to differences in curriculum between the school types and the possibility to revise initial tracking decisions after middle school (Mühlenweg and Puhani, 2010; Dustmann *et al.*, 2016).

The trend in the distribution of students in each school type across grades is shown in Appendix Table D.1. About 45 percent of 5^{th} graders in Hesse attend the academic-oriented higher school type. The share declines slightly as years in middle school increase. This is both because some students are downgrading to the non-academic school tracks and because there is an increase over time in the share of students who attend the higher school track. Of the remaining 55 percent who attend the non-academic school types, students are equally distributed across the three school types - the intermediate school, comprehensive school and the support stage. It is worth noting that the share of students who avoid tracking by attending the comprehensive schools is increasing over the years (from 16 percent to 19 percent between the 2007/08 and 2011/12 school years). On the other hand, the share of students in grade 9 that attends the lower middle school type has declined from 17 percent to 12 percent. This trend illustrates the

[2] Some states postpone middle school tracking until age 12 (end of grade 6) (Mühlenweg, 2007).

ongoing debate in Germany on the timing and the extent of tracking after elementary school. In most states, there is a push towards combining the lower and intermediate school tracks. Because only the higher school track leads to a degree allowing university/college entrance, our main outcome variables will be binary indicators of higher school track recommendation or choice. This effectively combines the intermediate and lower school tracks with comprehensive schools as non-tracking institutions.

Table 1 shows the descriptive statistics. Column (1) shows averages and standard deviations for the full sample whereas column (2) limits the sample to observations with valid data on teacher track recommendation, the track recommendation (TR) sample. The average age, gender and nationality are rather similar between the full and the TR sample. So are average class and enrollment sizes. Appendix Figure D.1 shows the distribution of class size and total enrollment. A typical classroom in 4^{th} grade consists of 20 students on average. This is comparable to the average class size in the US and most Western European countries. Only six percent of students are taught in classes that exceed the maximum class size of 25. The average enrollment size is about 60 with a standard deviation of 26. About six percent of students are enrolled in schools with enrollment size of more than 100.

About half of 4^{th} grade students receive a recommendation to attend the more academic school type. The school type recommendation is missing for about one third of students and the share of students who attend the higher school track is larger in the track recommendation sample than in the full sample (54 versus 45 percent).[3] There is a large overlap between track recommendation and track choice: in our sample, 42.7 and 48.3 percent of students follow the recommendation not to attend and to attend the higher school track, respectively. 5.3 percent of all students "upgrade" in the sense that they still attend the higher school track although that have been recommended not to attend it, whereas 3.6 percent of students "downgrade" in the sense that they do not attend the higher school track although they have been recommended to attend it. This means that the overlap of track recommendation and track choice is very high at 91 percent.

[3] Among the students with school track recommendation missing, the share of students who are not tracked after grade 4 is higher (60 percent) than among students whose track recommendation is not missing (24 percent), either because they decided to attend a comprehensive school (*Gesamtschule*, 30 compared to 13 percent), deferred the tracking decision by two years by entering the support stage (*Förderstufe*, 26 compared to 11 percent), or because they repeated the 4^{th} grade of elementary school (4 compared to 0 percent). Because all these choices are coded as 0 in the actual school track choice variable, where only "higher school track" is coded as 1, school track recommendation is not missing at random with respect to actual school track choice.

3 Empirical Strategy and Estimation Results

3.1 Identification Based on the Maximum Class Size Rule

To disentangle the effect of class size from other factors that influence educational outcomes and might be correlated with class size, our empirical strategy relies on maximum class size rules. The approach—pioneered by Angrist and Lavy (1999)—exploits the exogenous variation in actual class size arising from variation in enrollment size and the rule that governs the maximum number of students to be taught in one class. This rule creates a discontinuity in the relationship between actual class size and total enrollment. The assigned class size is obtained using the following formula:

$$ACS_{st} = E_{st}/[int(\frac{E_{st}-1}{25}) + 1] \tag{1}$$

where ACS_{st} is the assigned 4^{th} grade class size in school (s) at year (t), E_{st} is the total enrollment in school (s) at school year (t) and 25 is the official maximum class size for elementary schools in Hesse.

Figure 1 shows the first stage and reduced form relationships. Subfigure (a) shows the first stage relationship between the actual class size and the assigned class size generated using equation (1). To create the figures, the data are first collapsed at the classroom level and then averaged for each enrollment size. Subfigure (b) shows the relationship between the assigned class size and the higher school track attendance rate whereas subfigure (c) shows the relationship between the assigned class size and the average higher school track recommendation rate.

In subfigures (b) and (c), there is no visible similarity between the up-and-down pattern in assigned class size and the pattern in higher school track attendance or higher school track recommendation. Subfigure (a), however, shows similar up-and-down patterns between assigned class size and average class size as they depend on total enrollment. There is a jump in actual and assigned class size whenever total enrollment reaches integer multiples of the maximum class size rule. Schools with total enrollment just below the threshold have relatively larger classes compared to schools with total enrollment just above the threshold. The deviation of the actual class size from the assigned class size indicates that sometimes schools do not strictly follow the maximum class size rule. The assigned class size is, therefore, used as an instrument for the actual class size following a fuzzy regression discontinuity design (Hahn *et al.*, 2001).

Note that when we count how many schools during the five school years' observation period stick to the class size rule in at least one of the school years, the result is *all* 1,156 schools.

However, almost all schools, that is 947, in at least one of the five school years creates fewer classrooms than it should according to the class size rule, whereas only 244 schools in at least one school year create more classrooms than they should according to the rule. Given that teachers in Germany are civil servants, it is difficult to temporarily hire new teachers to form additional classrooms, which explains that almost all schools sometimes are not able to create additional classrooms.[4]

Although non-compliance does not necessarily harm our instrumental variables approach, our quasi-experimental identification strategy based on the maximum class size rule relies on the assumption that schools do not manipulate enrollment at the point of discontinuity. Figure 2 therefore shows the density of normalized enrollment (all cutoffs such as 25, 50, 75 *etc.* are defined to be 0 so that there should be larger and fewer classrooms with normalized enrollment up to 0 and smaller and more classrooms with normalized enrollment of 1 or higher as in Fredriksson *et al.* (2013). The figure, which also shows 95 percent confidence intervals, is generated using the Stata procedure provided by McCrary (2008) and limiting the sample to normalized enrollment values between -10 and +10. The test statistic for a jump in the density at the discontinuity is -0.027 with a standard error of 0.146, hence there is no evidence for a break in the density of enrollment at the point of discontinuity. What is striking, however, is that at *both* sides of the cut-off point, the density in enrollment near the point of discontinuity is smaller than it is further away from the cutoff. This implies that schools are designed in sizes such that they are less likely to be close to the point of discontinuity. This empirical finding is at least consistent with the regulations of the State of Hesse which state that classrooms should be built such that they can be continued in the following year. Schools close to the point of discontinuity have a harder time fulfilling that criterion or they have to deviate from the class size rule.[5] In sum, we assume that assigned class size is a valid instrument for actual class size and hence can be used to identify the causal effect of class size on educational outcomes. Note that the panel nature of our data also allows controlling for school fixed effects, so that the variation in the data which we use to identify the effect of class size on tracking outcomes comes from within school variation in enrollment size across the points of discontinuity over time.

[4] When classrooms are split, their size is almost, but not exactly, even. At the school by year level (counting each school in each year as a separate observation), there are 4,577 school by year observations in our full sample. 2,139 of these school by year observations have class size deviations of more than 1 from the average class size in grade 4. The number of school by year observations exhibiting a deviation from the average class size of more than 2 is 1,472. For deviations of more than 3, 4, and 5, the figures drop to 863, 469, and 203, respectively.

[5] See §1 of the regulation *Verordnung über die Festlegung der Anzahl und der Größe der Klassen, Gruppen und Kurse in allen Schulformen* which is available at `http://www.rv.hessenrecht.hessen.de/lexsoft/default/hessenrecht_rv.html?doc.hl=1&doc.id=hevr-AssBFSchulAPrVHE2011rahmen&documentnumber=1&numberofresults=1&showdoccase=1&doc.part=R&paramfromHL=true#docid:7885506,2,20170617`.

Our school fixed-effects fuzzy regression discontinuity design identification strategy is implemented by an instrumental variable approach. The first stage equation takes the following form:

$$CS_{st} = \alpha_0 + \alpha_1 ACS_{st} + \alpha_2 E_{st} + \gamma X_{ist} + \tau_s + \pi_t + \epsilon_{ist} \tag{2}$$

where CS_{st} is actual 4^{th} grade class size in school s during school year t and ACS_{st} is the assigned class size as defined in equation (1). E_{st} is a third-order polynomial or a piecewise linear trend of enrollment size and the coefficient α_2 captures any remaining continuous relationship between total enrollment and actual class size other than the discontinuous relationship captured through ACS_{st}. X_{ist} is a vector of student level background characteristics, τ_s are school fixed effects and π_t are year fixed effects.

3.2 Effects of Class Size on Educational Outcomes

We mainly use the following equation to estimate the effect of class size on educational outcomes:

$$Y_{ist} = \beta_0 + \beta_1 CS_{st} + \delta X_{ist} + \tau_s + \pi_t + \mu_{ist} \tag{3}$$

where Y_{ist} is the educational outcome of student i in school s during year t that is either the teacher's school type recommendation or the actual school type attended in 5^{th} grade. CS_{st} is the actual class size at the beginning of 4^{th} grade which is instrumented by the assigned class size as defined in equation (1). X_{ist} is a vector of controls such as students' gender, age and nationality. It also includes a third-order polynomial or a piecewise linear trend of enrollment size in order to capture any remaining continuous influence of enrollment size on tracking outcomes other than its discontinuous influence through the assigned class size. τ_s are school fixed effects which control for sorting of students between schools whereas π_t are year fixed effects.

Table 2 presents our main estimation results. We present two types of approaches. The results in panel A use predicted class size as in equation (2) as the instrument, the results in panel B use a binary instrument based on normalized enrollment such that the all cut-off points (25, 50, 75 *etc.*) are defined as zero and the instrument equals 1 when enrollment is above and 0 when it is below the cutoff, similar to Fredriksson *et al.* (2013). We then restrict the sample to enrollments that are 10 students away from the cut-off point at the maximum. In panel A, we estimate models with different specifications concerning the functional form of the running variable enrollment: the first three specifications use linear, second-order polynomial and third-order polynomial controls for enrollment. The fourth specification uses a piecewise

linear trend (spline, using Stata's *mkspline* command) instead of a polynomial. In panel B, we use a second-order polynomial and interact it with dummy variables for each segment with enrollments of at the maximum ±10 away from the points of discontinuity (25, 50, 75 *etc.*). The segment fixed effects and their interactions with the second-order polynomial in enrollment are taking account of the different slopes of enrollment around the various points of discontinuity as in Fredriksson *et al.* (2013). Column (1) presents models with track recommendation, column (2) with school choice as outcome variable for the full sample, and finally—as a robustness check—column (3) presents models with school choice as the outcome but only for the track recommendation sample, that is observations for which track recommendation is not missing.

When using predicted class size as an instrument, point estimates in columns (1) and (2) are small and not statistically significant (with only one exception with a first-order polynomial significant at the 10 percent level). In column (3) where we use school choice as the dependent variable but exclude observations with track recommendation missing from the sample all coefficients are statistically significant at the 10 percent level. The size of the estimate is mostly -0.003 which means that for an increase in class size of 10 students, the probability to attend the higher school track decreases by 3 percentage points. Given that 45 percent of students attend the higher track, this effect is not very large but still worth mentioning. When using the approach by Fredriksson *et al.* (2013) (panel B), none of the estimates are statistically significant and point estimates are close to zero.

Table 3 displays the first-stage regression coefficients with the corresponding F-statistics for the estimates reported in Table 2. Columns (1) and (3) of the table show that assigned class size is a strong predictor of actual class size. The coefficients in panel A are positive and significant at the 1 percent level. The F-statistics in columns (2) and (4) are well above the rule-of-thumb threshold for a weak instrument of 10 (Staiger and Stock, 1997). The first-stage coefficients for the approach by Fredriksson *et al.* (2013) are also statistically significant at the 1 percent level, but F-statistics are smaller than in panel A, because the approach by Fredriksson *et al.* (2013) is more demanding of the data.[6] This is also reflected in the somewhat larger standard errors associated with the estimates in panel B than in panel A of Table 2. For the subsequent robustness checks and heterogeneous effects, we will therefore stick to the specifications of panel A.

[6] Note that the first-stage coefficients in panel B are negative because the binary instrument is defined such that observations above the cutoff point—for which classes are smaller—are coded as 1 and those below the cutoff point are coded as 0.

3.3 Robustness Checks and Heterogeneous Effects

As mentioned in Section 2, one of the unique features of the Hessian administrative data is that it contains a time-consistent student identifier which makes it possible to follow students as they increase their level of schooling. So far, we have used the panel dimension of the administrative data in order to relate enrollments and class sizes at the end of elementary school (4^{th} grade) to the subsequent educational outcomes at the beginning of middle school (5^{th} grade). The administrative data allows us to follow two school cohorts from 1^{st} grade until they enter different types of middle school. By restricting the sample to these two school cohorts, we are able to generate an instrument for class size based on total enrollment in grade 1 instead of grade 4. This may be important because enrollment size in grade 4 could be endogenous due to, for instance, grade repetition in lower grades. In addition, schools are more likely to implement the maximum class size rule when students start school in 1^{st} grade.

In Table 4, we therefore exploit the panel nature of our data and use enrollment in grade 1 to build our instrument of assigned class size. In the table, we use assigned class size in grade 1 as an instrument for the impact variable class size in grade 1. Columns (1) to (3) use the same outcome variables as in Table 2: (1) track recommendation, (2) track choice for the full sample, and (3) track choice for the sample with valid information on track recommendation. As can be seen from the table, none of the coefficients of class size in columns (1) to (3) are statistically significant. In column (4), where grade repetition is the outcome variable, coefficients are statistically significant at the 1 percent level across all specifications. The size of the estimate, which is always 0.004, implies that 10 more students in a grade 1 classroom increase the probability to experience a grade repetition in elementary school by 4 percentage points. Given that only 6.6 percent of students repeat a grade any time from 1^{st} to 4^{th} grade, this is a large effect. However, as we can only use data for two school cohorts to estimate models relating class size in grade 1 to tracking outcomes in grade 5, our school fixed-effects models in Table 4 are similar to first-difference models. It would be preferable to have a longer panel of data to check the robustness of this large—that is large in relation to the mean of the dependent variable—effect of class size on grade repetition. If this effect were confirmed in a future study with a longer panel, it would probably be the most remarkable effect of class size in our study.

Given that overall we find either no effects of class size on track recommendation or track choice or effects that are not robust, we provide estimates by subgroups in Table 5. Previous studies find a larger benefit of smaller classes for students from a disadvantaged background where disadvantaged background is defined, for instance, based on parental education, family

size, number of books at home or location of schools. Unfortunately, the Hessian administrative LUSD data does not contain information on direct measures of disadvantaged family background. In order to test for heterogeneous class-size effects, we rely on information on students' nationality or gender. In Table 5, we go back to our initial specification with enrollment in grade 4 as the instrument and class size in grade 4 as the impact variable. The upper half of the table splits the sample by gender, the lower half of the table provides separate estimates for non-European and European students. In both cases enrollment and class size are defined on the total student population in grade 4 before the sample is split. Some interesting results emerge from the table. Across all specifications, male students seem slightly harmed by larger classes. The coefficient for track recommendation is -0.005 in most specifications and statistically significant at the 5 percent level. This implies that 10 more students in a classroom decrease the probability of being recommended for the higher school track by 5 percentage points for male students. Given that about half of the students are recommended for the higher track, the effect amounts to about 10 percent of the mean of the dependent variable. This effect is not huge given the large class size change of 10 students but still worth considering. When track choice is the outcome variable, the estimate is -0.003 and statistically significant at the 10 percent level. This means that a class size increase of 10 students decreases the probability of choosing the higher school track by 3 percentage points. Note that in the "track recommendation sample", the effect is twice as large at -0.006 and statistically significant at the 1 percent level.

When splitting the sample by European versus non-European citizenship, we find no significant effects of class size on either track recommendation or track choice for non-European students, with the exception of track choice in the track recommendation sample, where the point estimate is mostly -0.013 and statistically significant at the 10 percent level. This large point estimate, implying that 10 more students reduce higher track choice by 13 percentage points suggests that deleting observations with track recommendation missing—as artificially done in the track recommendation sample—may lead to bias of the track choice coefficient, especially when specific subgroups are considered.[7] In the sample of students with European citizenship, the effect of class size on track recommendation is estimated to be -0-003 and statistically significant at the 10 percent level, whereas the effect on track choice is close to zero at -0.001 and not statistically different from zero in the full sample. In the track recommendation sample, the estimate is somewhat larger at -0.003 and mostly statistically significant at the 10 percent level. Again, this suggests that restricting the observations to the track recommendation

[7] Note that in previous estimates—which relied on the full sample or larger subsamples such as all male students—the differences between the full and the track recommendation samples were not so large.

sample leads to a slightly negative bias of the estimation coefficient. Still the estimates for the European sample are not very large suggesting that the effect of an increase of 10 students in class size reduces the probability of receiving a recommendation for the higher track by about 3 percentage points.[8]

4 Conclusion

A student-level administrative data set from the German state of Hesse allowed us to rather precisely estimate the causal effect of class size in elementary school on middle school tracking recommendation and on school choice. Exploiting the exogenous variation in class size arising from maximum class size rules, despite of using "big data", we mostly find no or only small effects of class size on receiving a recommendation for or on the actual choice of attending the more academic middle school type. For male students, we find that an increase in class size by 10 students reduces their chance of attending the higher school track—which more than 40 percent of students attend—by 3 percentage points. We also find that increasing class size by 10 students in grade 1 of elementary school increases the probability for any grade repetition in elementary school by 4 percentage points. This effect is large given that only 6.6 percent of students in our sample repeat a grade in elementary school, but it is based on only two school cohorts of 1^{st} graders in a school fixed effects model. A panel with a larger time dimension should be available in the future to check the robustness of this result.

The potential explanation for the mostly insignificant effect of class size on tracking which we find in this paper could be that the average size of classes in the German state of Hesse is quite small and hence it might be below the threshold that is relevant for students' academic achievement. In Israel the threshold for splitting a classroom is 40 students (Angrist and Lavy, 1999), in Sweden it is 30 students (Fredriksson *et al.*, 2013), whereas it is only 25 students in the German state of Hesse. Given that these authors find more positive effects of smaller classrooms, our findings point to the importance of what we mean by "large" and "small" classrooms. The other explanation is related to the educational outcome measures available in our administrative data. The type of middle school students attend or are recommended to attend could be imprecise measures of academic achievement compared to test scores or academic

[8] We have also experimented with splitting the sample into students with above and below average shares of females and non-Europeans in the classroom. The main results were not statistically significant. Only when restricting the sample with above average share of females in the classroom to the track recommendation sample for the outcome school choice, coefficients were mostly -0.004 and statistically significant at the 10 percent level. Again, we interpret this as a slightly negative bias associated with the track recommendation sample.

grades. This calls for further research that matches the administrative data we use in this paper with survey or other administrative data sources which contain precise measures of students' academic achievement along with richer information on family background characteristics. It would also be important to provide data which make it possible to study long-term effects.

A Acknowledgements

This research would not have been possible without the onsite data access provided by the Ministry of Culture and Education of the State of Hesse (*Hessisches Kultusministerium*) in cooperation with the Research Data Center (*Forschungsdatenzentrum*) of the Statistical Office of the State of Hesse (*Hessisches Statistisches Landesamt*). We thank anonymous referees, Manuel Boos, Marc Deutschmann, Peter Gottfried, Nicole Purnhagen, Alexander Richter, and Claudia Schäfer-Sold for helpful comments. All remaining errors are our own. The views expressed in this article are those of the authors and do not necessarily reflect the views of the Ministry of Culture and Education or the Research Data Center. While working on this paper, Bethlehem Argaw was a researcher in the Labour Markets, Human Resources and Social Policy Department of the Centre for European Economic Research (ZEW).

B References

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

Bonesroenning, H. (2003). Class size effects on student achievement in Norway: Patterns and explanations. *Southern Economic Journal*, 69(4), 952–965.

Browning, M. and Heinesen, E. (2007). Class size, teacher hours and educational attainment. *Scandinavian Journal of Economics*, 109(2), 415–438.

Card, D. and Krueger, A. B. (1996). School resources and student outcomes: An overview of the literature and new evidence from North and South Carolina. *Journal of Economic Perspectives*, 10(4), 31–50.

Ding, W. and Lehrer, S. F. (2010). Estimating treatment effects from contaminated multiperiod education experiments: The dynamics impacts of class size reductions. *Review of Economics and Statistics*, 92(1), 31–42.

Dobbelsteen, S., Levin, J., and Oosterbeek, H. (2002). The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition. *Oxford Bulletin of Economic and Statistics*, 64(17), 19–38.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2016). The long-term effects of early track choice. *Economic Journal*, 127(603), 1348–1380.

Dustmann, C., Rajah, N., and van Soest, A. (2003). Class size, education, and wages. *Economic Journal*, 113(485), F99–F120.

Fredriksson, P., Oeckert, B., and Oosterbeek, H. (2013). Long-term effects of class size. *The Quarterly Journal of Economics*, 128(1), 249–285.

Gary-Bobo, R. J. and Mahjoub, M. B. (2006). Estimation of class-size effects using 'Miamonides' Rule': The case of French junior high schools. *CEPR Discussion Paper No. 5754, London.*

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a Regression-Discontinuity Design. *Econometrica*, 69(1), 201–209.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, 113(485), F64–F98.

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

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

Krueger, A. B. (2003). Economic considerations and class size. *Economic Journal*, 113(485), F34–F63.

Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project Star. *Economic Journal*, 111(468), 1–28.

Leuven, E., Oosterbeek, H., and Roenning, M. (2008). Quasi-experimental estimates of the effect of class size on achievement in Norway. *Scandinavian Journal of Economics*, 110(4), 663–693.

Levin, J. (2001). For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement. *Empirical Economics*, 26(1), 221–246.

Mühlenweg, A. M. (2007). *Educational effects of early or later secondary school tracking in Germany.* ZEW-Centre for European Economic Research Discussion Paper 07-079, Mannheim.

Mühlenweg, A. M. and Puhani, P. A. (2010). The evolution of the school-entry age effect in a school tracking system. *Journal of Human Resources*, 45(2), 407–438.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698–714.

Piketty, T. (2004). *Should we reduce class size or school segregation? Theory and evidence from France.* Presentation at the Roy Seminars, Association pour le développement de la recherche en économie et en statistique (ADRES), available at: http://adres.ens.fr/IMG/pdf/22112004b.pdf, accessed on December 02, 2017.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3), 557–586.

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

Wößmann, L. (2005). Educational production in Europe. *Economic Policy*, 20(43), 445–504.

Wößmann, L. and West, M. R. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review*, 50(3), 695–736.

C Tables and Figures

Table 1: Sample summary statistics

	Full Sample	TR Sample
Educational outcomes		
Higher track recommendation	-	0.52
Higher track attendance	0.45	0.54
Student characteristics		
Age	9.86 (0.50)	9.84 (0.49)
Male	0.51	0.51
Non-European nationals	0.07	0.07
School characteristics		
Class size	20.37 (3.82)	20.45 (3.71)
Enrollment size	59.69 (26.49)	60.10 (26.67)
Number of school years	5	5
Number of schools	1,156	1,101
Number of classes	13,317	10,482
Number of students	258,098	168,063

Note: Standard deviations are shown in brackets when applicable. TR (Track Recommendation) sample refers to observations with valid data on school track recommendation.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*).

Table 2: Main estimation results: Coefficients of class size

Enrollment Specification	Track Recom.	Track choice - full sample	Track choice - TR sample
	(1)	(2)	(3)
A. Predicted class size as an instrument			
1^{st} order polynomial	-0.003*	-0.001	-0.004**
	(0.002)	(0.001)	(0.002)
2^{nd} order polynomial	-0.003	-0.001	-0.003**
	(0.002)	(0.001)	(0.002)
3^{rd} order polynomial	-0.003	-0.001	-0.003*
	(0.002)	(0.001)	(0.002))
Piecewise linear trend	-0.003	-0.001	-0.003*
	(0.002)	(0.001)	(0.002)
Number of schools	1,101	1,156	1,101
Number of students	168,063	258,098	168,063
B. Binary instrument based on normalized enrollment as in Fredriksson *et al.* (2013)			
Segment fixed effects, 2^{nd} order enrollment polynomial	-0.002	0.001	-0.001
and their interactions	(0.004)	(0.003)	(0.005)
Number of schools	1,026	1,094	1,026
Number of students	129,976	197,845	129,976

Note: Instrumental variable estimates where class size in grade 4 is instrumented with assigned class size in grade 4. Standard errors are clustered at the school level and shown in parentheses. TR sample refers to observations with valid data on track recommendation. All regressions include control for age, gender, an indicator for being non-European national, interaction between gender and non-European national, year dummies and school fixed effects. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schüler-datenbank, LUSD*).

Table 3: First stage results of the main estimates: Coefficients of instrument

Enrollment Specification	TR sample	*F*-stat.	Full sample	*F*-stat.
	(1)	(2)	(3)	(4)
A. Predicted class size as an instrument				
1^{st} order polynomial	0.501***	403.01	0.514***	619.22
	(0.025)		(0.021)	
2^{nd} order polynomial	0.491***	373.94	0.503***	567.00
	(0.025)		(0.021)	
3^{rd} order polynomial	0.487***	358.17	0.498***	541.27
	(0.026)		(0.021)	
Piecewise linear trend	0.481***	343.71	0.491***	516.70
	(0.026)		(0.022)	
Number of students	168,063		258,098	
B. Binary instrument based on normalized enrollment as in Fredriksson *et al.* (2013)				
Segment fixed effects, 2^{nd} order enrollment polynomial	-2.185***	55.26	-2.265 ***	85.88
and their interactions	(0.294)		(0.244)	
Number of students	129,976		197,845	

Note: Standard errors are clustered at the school level and shown in parentheses. TR sample refers to observations with valid data on school recommendation. All regressions include control for age, gender, an indicator for being non-European national, interaction between gender and non-European national, year dummies and school fixed effects. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schüler-datenbank, LUSD*).

Table 4: Coefficients of class size in first grade instrumented by assigned class size in first grade

Enrollment Specification	Track Recommendation	Track choice - full sample	Track choice - TR sample	Any grade repetition (Grade 1 - 4)
	(1)	(2)	(3)	(4)
1^{st} order polynomial	-0.001	-0.003	-0.003	0.004***
	(0.004)	(0.002)	(0.004)	(0.001)
2^{nd} order polynomial	-0.001	-0.003	-0.002	0.004***
	(0.004)	(0.002)	(0.004)	(0.001)
3^{rd} order polynomial	-0.001	-0.003	-0.003	0.004***
	(0.004)	(0.002)	(0.004)	(0.001)
Piecewise linear trend	-0.001	-0.004	-0.002	0.004***
	(0.004)	(0.003)	(0.004)	(0.001)
Number of students	56,809	93,271	56,809	99,609

Note: Instrumental variable estimates where class size in grade 1 is instrumented with assigned class size in grade 1. Standard errors are clustered at the school level and shown in parentheses. TR sample refers to observations with valid data on school recommendation. All regressions include control for age, gender, an indicator for being non-European national, interaction between gender and non-European national, year dummies and school fixed effects. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schüler-datenbank, LUSD*).

Table 5: Differences across groups: Coefficients of class size

Enrollment Specification	Track Recom.	Track choice - full sample	Track choice - TR sample	Track Recom.	Track choice - full sample	Track choice - TR sample
	Female Sample			Male Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
1^{st} order polynomial	-0.001	0.001	-0.011	-0.005**	-0.003**	-0.006***
	(0.002)	(0.001)	(0.132)	(0.002)	(0.001)	(0.002)
2^{nd} order polynomial	-0.001	0.001	-0.001	-0.004 **	-0.003*	-0.006***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
3^{rd} order polynomial	-0.001	0.001	-0.001	-0.005**	-0.003*	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Piecewise linear trend	-0.001	0.001	-0.001	-0.005**	-0.003*	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of students	83,070	126,783	83,070	84,969	131,314	84,969
	Non-European Sample			European Sample		
	(7)	(8)	(9)	(10)	(11)	(12)
1^{st} order polynomial	-0.004	-0.002	-0.013*	-0.003*	-0.001	-0.004**
	(0.007)	(0.005)	(0.007)	(0.002)	(0.001)	(0.002)
2^{nd} order polynomial	-0.004	-0.002	-0.013*	-0.003*	-0.001	-0.003**
	(0.007)	(0.005)	(0.007)	(0.002)	(0.001)	(0.002)
3^{rd} order polynomial	-0.004	-0.002	-0.013*	-0.003*	-0.001	-0.003*
	(0.007)	(0.005)	(0.007)	(0.002)	(0.001)	(0.002)
Piecewise linear trend	0.005	-0.002	-0.015**	-0.003*	-0.001	-0.003*
	(0.007)	(0.005)	(0.007)	(0.002)	(0.001)	(0.002)
Number of students	11,639	18,934	11,639	156,302	239,049	156,302

Note: Instrumental variable estimates where class size in grade 4 is instrumented with assigned class size in grade 4. Standard errors are clustered at the school level and shown in parentheses. TR sample refers to observations with valid data on school recommendation. All regressions include control for age, gender, an indicator for being non-European national, interaction between gender and non-European national, year dummies and school fixed effects. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schüler-datenbank, LUSD*).

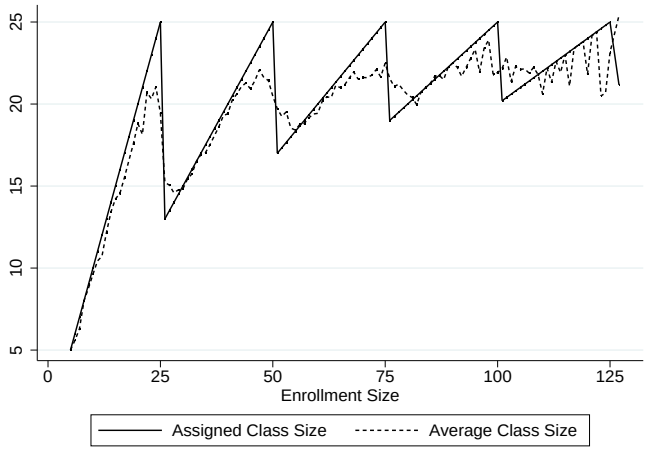
D Appendix

Table D.1: Distribution of students across grades and school types

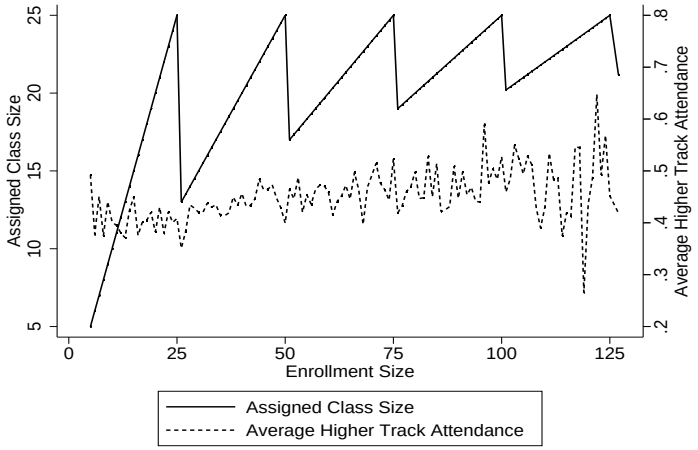
	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
School year 2007/08					
Higher school type	0.44	0.42	0.42	0.40	0.37
Non-higher school types					
Intermediate school	0.17	0.16	0.28	0.29	0.30
Lower track school	0.04	0.05	0.14	0.16	0.17
Comprehensive school	0.18	0.18	0.17	0.16	0.16
Support stage	0.18	0.19	-	-	-
Number of students	59,264	57,259	56,940	59,659	60,570
School year 2011/12					
Higher school type	0.46	0.44	0.44	0.42	0.40
Non-higher school types					
Intermediate school	0.16	0.16	0.26	0.28	0.29
Lower track school	0.03	0.04	0.10	0.12	0.12
Comprehensive school	0.20	0.19	0.20	0.19	0.19
Support stage	0.16	0.16	-	-	-
Number of students	53,322	56,221	58,290	60,300	62,310

Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*).

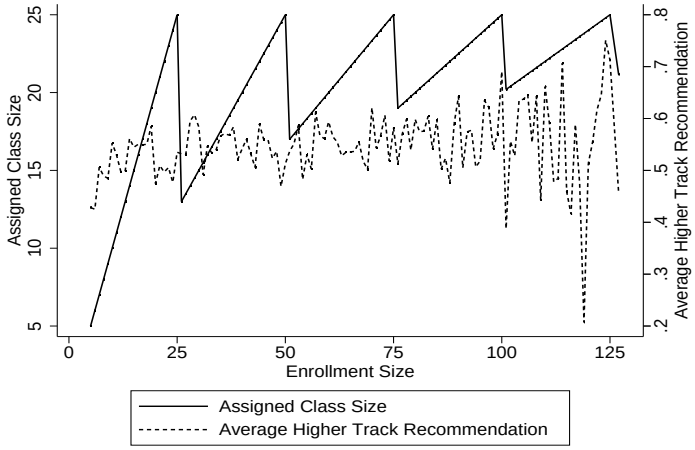
Figure 1: First stage and reduced form relationships



(a) Assigned class size and actual class size by total enrollment.



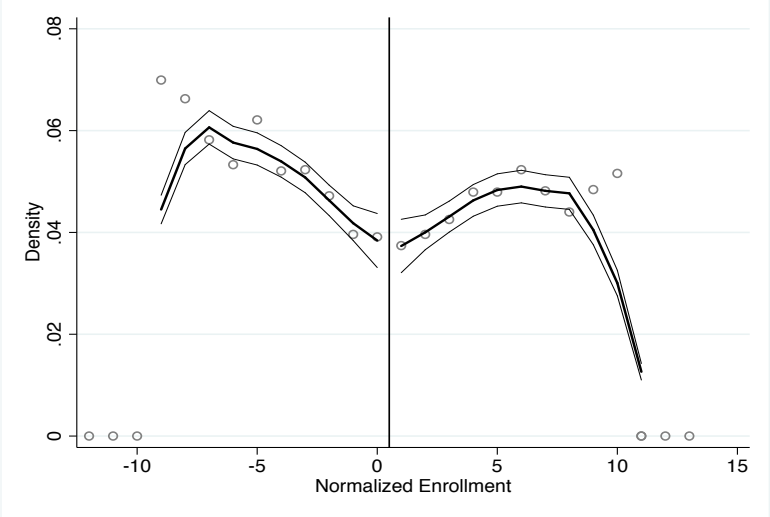
(b) Assigned class size and average higher school track attendance by total enrollment.



(c) Assigned class size and average higher school track recommendation by total enrollment.

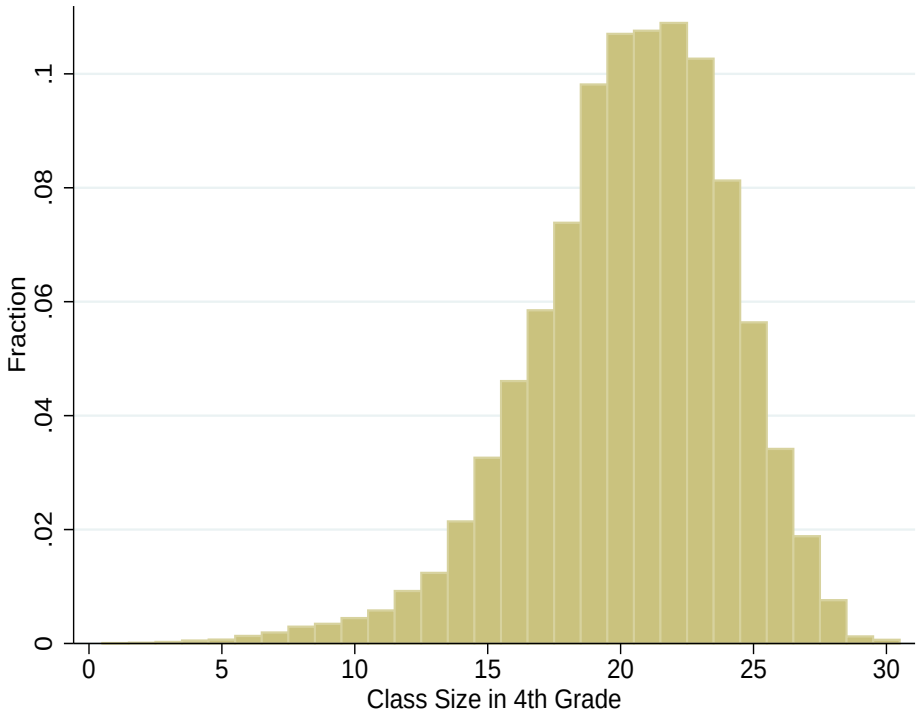
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*).

Figure 2: McCary density test

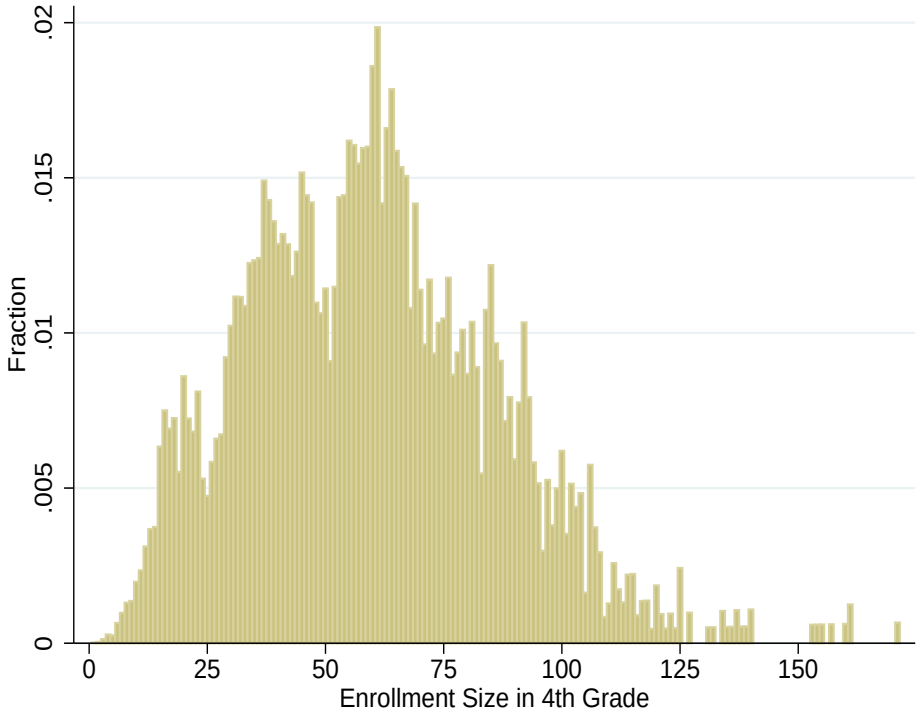


Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*).

Figure D.1: Distribution of class size and total enrollment



(a) Distribution of class size.



(b) Distribution of total enrollment.

Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*).