Peer Effects and Academic Achievement

Regression Discontinuity Approach

Arna Vardardottir*

March 2, 2012

Preliminary draft

- please do not cite

Abstract

In this paper, I study ability peer effects in an Icelandic high school. The identification relies on a fuzzy regression discontinuity approach where student assignment into high-ability classes constitutes the source of identifying information. An important feature of this system is that the same teachers teach high-ability and normal classes, both types of classes follow a common curriculum and all students take the same exams. Furthermore, the system is unofficial so students are in most cases not aware of it before they have started their studies. In cases where they are aware of the system's existence they do not know where the threshold lies prior to enrolment and they are unlikely to be able to attend other high-schools if they decide to drop out once they learn whether they have been assigned to a high-ability class. I find that sorting students into high-ability classes does have significant and sizable effect on the academic achievement of students around the assignment threshold, i.e., the results suggest that being assigned to a high-ability class increases academic achievement by 0.23 standard deviations.

*Department of Economics, Stockholm School of Economics, SE-113 83 Stockholm, Sweden. E-mail: arna.vardardottir@hhs.se

1 Introduction

Peer effects in education are generally accepted to be of importance. Despite this belief there is no general consensus on the direction of the effect peers have on one another. Different theories attempt to explain this and according to some of them the average ability of classmates has detrimental effect on one's schooling outcomes while others imply that it enhances ones achievements (Marsh, 2005). Furthermore, the exact causal mechanism of peer effects in education is also ambiguous. One possible, and most direct, channel for peer effects is that students instruct each other. Other possible channels are for instance classroom disruption and classroom atmosphere. Students could also be indirectly affected by their peers. This can for instance come about through the way teachers react to different groups of students. Another possibility is if students are sorted into classes based on their ability it might allow teachers to match instructions more closely to students needs because of more homogenous group, which would benefit all students. However, my primary purpose with this paper is to establish empirically the existence and direction of peer effects but not to distinguish the channels by which peer effects operate.

In recent years the estimation of peer effects in schools has received much attention. Several studies have provided important findings about these effects in different circumstances. Among the studies finding that students benefit from being around high-achieving peers are Hoxby (2000), Sacerdote (2001), Zimmerman (2003), McEwan (2003), Groux and Maurin (2006), Hoxby and Weingarth (2006), Graham (2008), and Ammermueller and Pischke (2009).

In this paper I examine the effects of the ability of a student's classmates on her academic success in her first year of college, i.e., I explore whether better academic qualifications of a student's classmates can cause an effect on achievement. The problem when it comes to estimating peer effect is that, as the saying goes, birds of feather flock together[1], and the same applies to students. College students self-select their friends and they are likely to select friends whose unobservable characteristics are systematically related to theirs. Even when students do

[1]The saying describes the tendency of individuals to associate with others who are similar to themselves, a phenomenon known as homophily.

not select their peers entirely voluntary there can be a relation between their characteristics. If a student decides, for instance, to enroll in a demanding course that is non-compulsory his classmates will probably have similar characteristics. Most high-school and college students choose their peers and therefore it is difficult to estimate peer-effects in most higher-education settings. In situations where students choose their own peers we are subject to the reflection problem, i.e., if a student's peers have unobserved characteristics that are systematically related to her own, estimation of peer effects cannot be given a causal interpretation. If, for instance, a smart student tends to choose smart peers then it is not feasible to statistically distinguish between the effects of the students intelligence and the effect of peers intelligence. In this paper I address this problem by employing a regression discontinuity (RD) design where student assignment into high-ability (HA) classes constitutes the source of identifying information. The basic intuition behind this approach is that, in the absence of program manipulation, students just below the treatment-determining grade cutoff should provide valid counterfactual outcomes for students just above the cutoff, who were assigned to HA classes.

I use data on 5 years of entering students at an Icelandic high school to test for peer effects among classmates. The outcome variable of interest is the academic achievement measured by the end of first year. There are approximately 270 incoming students each year that are divided into 10 classes out of which 3-4 classes are HA classes but the rest of the students are randomly assigned to normal classes. The system is not official so prior to enrolment students and their parents are, in most cases, not aware of the fact that streaming into classes will take place. Furthermore, once students learn wheather they are in a normal or HA class their outside options are rather limited if they decide to drop out. They will most likely have to wait at least one semester to get into another one and for a whole year in order to get into the most sought-after high-school. Also, the school under consideration has always been among the most-sought after schools in Iceland and it has always been considered to be a good signal of high-academic ability to have graduaded from there. Since it will be difficult for students to get as good a signal of their academic ability at another high-school at this point in time it is difficult to see why students would be willing to drop out and thereby let go of this signal because they did not get into a HA class. The same teachers teach normal classes and HA classes, they cover the same material and all students take the same exams. Selection into classes is mostly based on students' assignment grades, defined as the average of their results in Mathematics, Icelandic, English and Danish on the standardized tests for 10^{th} grade and their school grades in these subjects. The probability of being assigned to a HA class therefore jumps at the 60^{th} or 70^{th} percentile of the assignment grade, depending on which year we consider, and therefore I use a fuzzy RD design to test for peer effects, i.e., whether being assigned to a HA will affect one's grades.

Using a RD approach I restrict the estimation to the discontinuity in the assignment probability for a HA class since this will essentially result in a randomized experiment, i.e., I compare outcomes for the students whose grades are just below and just above this 60^{th} or 70^{th} percentile threshold since they on average will have similar characteristics except for the treatment. Students just above and just below the threshold are therefore treated as the treatment and control group, respectively. Those students slightly below the threshold provide the counterfactual outcomes for the students slightly above since the treatment status is randomized in a neighborhood of the threshold. Jumps in the relationship between assignment grade and grades by the end of the first year in the neighborhood of the HA class threshold can therefore be taken as evidence of a treatment effect. Due to the drawback of the discontinuity approach that there are usually few observations around the discontinuity most researchers apply the control function approach in practice. I will follow the same strategy in this paper.

The contribution of this paper is twofold. First, the way I measure peer ability is an improvement over existing studies. The majority of previous empirical evidence on ability peer effects in education comes from studies that are either based on data that does not include class identifiers or they examine the effect of academic ability of peers without having direct measures of their academic ability but rely instead on background characteristics as proxies for this. Since students spend a relatively big part of their time in class their classmates are very likely to be significantly influenced by their classmates. It is therefore very important to be able to identify this group. To the best of my knowledge, this is the first paper that is both able to identify classmates and measure peer ability directly using their test scores from national and school exams at the end of the 10^{th} grade. Second, I am not aware of any other study that pursues a fuzzy regression discontinuity strategy to extract the causal impacts of peers ability on achievement. However, Bui, Craig and Imberman (2011) have applied the same method to estimate the effects of gifted and talented services on students, i.e., they estimate combined effect of better peers, higher quality teachers and a change in curriculum. The big advantage of the setting of this paper over their setting is how clean the treatment is, the only difference between the normal and HA classes is the quality of peers since they are taught by the same teachers, they follow a common curriculum and take the same exams.

I find that assigning students to a class with students that are on average of higher ability in comparison to a class where students are on average of lower ability has a positive and significant effect on their academic performance. Specifically, my results suggest that increasing academic ability of peers by one standard deviation increases one's own academic performance by approximately 0.42 standard deviations. Visual results also provide evidence that academic achievement, as measured by spring exam results or year grade results, is affected by being assigned into a HA class and therefore fit with the estimates obtained.

The rest of the paper is organized as follows. In the next section I describe the institutional background and the dataset. Section 3 describes the identification strategy and discusses problems that come up when measuring the causal peer effect. Section 4 reports the main results while section 5 presents concluding remarks.

2 Previous Literature

When identifying the causal effect of peer ability on educational outcomes two issues are particularly challenging. First, students self-select their friends and they are likely to select friends whose unobservable characteristics are systematically related to theirs, i.e., the ability of peers is not exogenous to one's own ability and characteristics. If all observable and unobservable factors that determine educational achievements and individual sorting are not accounted for this will result in biased estimates of classroom peer effects. Second, it is difficult to identify the reason for why students who belong to the same group tend to behave similarly. In a pioneering study, Manski (1993) distinguishes between endogenous effects, correlated effects and exogenous (or contextual) effects that all could explain this phenomenon. It could be that similar behavior can be explained by endogenous effects, wherein the propensity of a student to do well varies with the prevalence of high academic achievement in the group. Similar behavior within groups could also stem from correlated effects, wherein individuals in the same group tend to behave similarly because they face similar environments and have similar personal characteristics. Lastly, the reason for similar group behavior could be exogenous effects, wherein individuals in the same group tend to behave similarly because of exogenous characteristics to the group.

One remedy is to randomly assign students to peer groups or assigning students into groups based only on measurable characteristics that can serve as controls in estimation. In recent years several studies have exploited random assignment to groups to overcome the reflection problem and identify the causal effect of peers' ability. For example, Sacerdote (2001) and Zimmerman (2003) present evidence on ability peer effects in college based on randomly paired roommates in university housing at Dartmouth College and Williams College, respectively. Sanbonmatsu et al. (2004) exploit a randomized housing mobility experiment in the US to identify the effect of neighbouhood characteristics on student school outcomes. They find that being given the option to move to a better neighborhood had zero and insignificant effect on students' educational performance. Graham (2008) exploits the random assignment in the STAR class size experiment in Tennessee to examine the effect of peer quality on kindergarten achievement. He pursues an identification strategy based on the fact that peer quality variance is greater across the subset of small classes than it is across larger ones and class type therefore provides a plausible source of exogenous variation in peer quality variance. However, random class assignment is not that common in higher education, so using this method to test for ability peer effects is seldom feasible and researchers must therefore resort to other methods to identify a causal effect of peers' ability in observational studies.

Other approaches are certainly on offer. Hoxby (2000) exploits exogenous variation in peer composition in adjacent years at the school grade-level in elementary schools in Texas. McEwan (2003) studies peer effects among eighth graders in Chile using a school fixed effect approach. Hanushek et al. (2003) rely on a student and school-by-grade fixed effects strategy and uses previous peer achievement as a measure of peer-group ability in order to eliminate the problem of simultaneity. Ammermueller and Pischke (2009) investigate peer effects in primary schools in several European countries, including Iceland, by employing a school fixed effect strategy and use the number of books at home as their peer group measure. Lavy, Silva and Weinhardt (2009) pursue an alternative identification strategy and analyze whether there is systematic correlation between variation in subject outcomes for a student and the variation in subject ability of his peers. Schindler Rangvid (2007) uses the Danish subsample of the PISA data to analyze school composition effect on student outcomes and employs a comprehensive set of controls from Danish register data in order to control for endogeneity in school choice. Schneeweis and Winter-Ebner (2007) use the Austrian subsample of the PISA data and employ school type fixed effects and school fixed effects to estimate peer effects. Groux and Maurin (2007) investigate whether teenagers in France are influenced by their neighbors relying on an instrumental variable approach, using neighbors' dates of birth to identify the effect of neighbors' early educational advancement on an adolescent's performance at school. Duflo, Dupas and Kremer (2008) use experimental data from Kenya where schools are randomly assigned to being tracking or non-tracking. They find positive effects on the academic achievement of students who were randomly assigned to academically stronger peers in the non-tracking schools. Compared to students in non-tracking schools, students in tracking schools scored substantially higher on exams after 18 months. A reasonable interpretation would be that there is a positive direct effect of peers quality and also an indirect effect, operating through teacher behavior. However, this context that may have limited applicability for education systems in developed countries. The indirect effect, stemming from the fact that teachers are able to teach at a level more appropriate to the average student, will very likely not be the same in developed countries where student heterogeneity is not as great.

The previous literature finds peer effects in education ranging from close to zero (Sanbonmatsu et al., 2004), to about 0.50 standard deviations (Hoxby, 2000; Boozer and Cacciola, 2001). In studies where it was possible to identify classmates, peer effects were found to be of somewhat greater magnitude than those who could only identify peers by school-grade, suggesting that studies do not identify classmates are possibly missing out on information on the "real" reference group of a student. The critical point in measuring the influence of peers is to identify the "real" peers. Keeping in mind that students spend a relatively big part of their time in class it seems to be a credible assumption that their classmates are a good proxy of their group of peers. However, in some cases there can be significant variation between classes within school-grades and hence the assumption that school grade peers are a good proxy of classmates can be quite strong.

3 The dataset

I use data on 5 years of entering students at the Commercial College of Iceland in Reykjavik to test for ability peer effects among classmates. Compulsory education in Iceland is organized in a single structure system, i.e., primary and lower secondary education belong to the same school level, and generally take place in the same school. The law concerning compulsory education stipulates that education shall be mandatory for children and adolescents between the ages of six and sixteen. Upper secondary education is not compulsory, but anyone who has completed compulsory education has the right to enter a course of studies in an upper secondary school. Students are usually 16-20 years of age. General academic education is primarily organized as a four-year course leading to a matriculation examination ('stúdentspróf'). At the time under consideration, students had to take standardized exams in Mathematics, Icelandic, English and Danish by the end of the 10^{th} grade in order to get into an upper secondary school. The grades from these exams and their grades from their primary school determined into which school they got.

The data set consists of 1353 students, 644 female and 709 male. The Commercial College of Iceland is a four-year senior high school / college for students who have completed the Icelandic compulsory education. In their first year all students follow a common curriculum. At the end of their second year students receive the Commercial Diploma which corresponds roughly to A-levels in the United Kingdom and the High School Diploma in the United States. During the remaining two years of their four-year program, students complete their matriculation examination. These two years could be considered comparable to two years of study at an academic college, for example equivalent to two years of university-level foundation courses in an American junior college.

Selection into classes is mainly based on students' assignment grades, defined as the average of their results in Mathematics, Icelandic, English and Danish on the standardized tests for students in the 10^{th} grade and their school grades in these subjects. There are approximately 270 incoming students each year and they are assigned to 10 different classes, where each class spends the entire school day together. Students are assigned to 3-4 HA classes (depending on the year) or to normal classes. Students above the 60^{th} or 70^{th} percentile threshold (depending on which year we consider) are much more likely to end up in HA classes than those below it. Students are randomly assigned into classes within each class-type. The only difference between being assigned to a normal class and a HA class is therefore that HA classes have peers of higher academic ability. In particular, the same teachers teach normal classes and HA classes. Specifically, each year every teacher teaches usually a couple of classes within each grade a specific subject and school authorites make sure that teachers that teach HA classes also teach normal classes. Furthermore, all classes cover the same material and they take the same exams. However, I do not have data on teachers so I cannot show it explicitly that both types of classes are taught by the same teacher. The reason why the assignment threshold differs by year is that in 1995-1997 there were 3 HA classes out of 10 classes in total and in 1998-1999 there were 4 HA classes out of 10. The outcome variable of interest is students' academic achievement of which I have 2 measures. The first is the normalized average grade from all the spring exams. The second measure is their normalized year grade that is based on all grades on hand-in assignments, quizzes and Christmas exams. In addition I have information on from which school students come, in which neighborhood they live and their year of birth. Tables (1) and (2) show the assignment grades and normalized assignment grades, respectively, for all the classes. Table (3) then shows the number of students in each class, where classes within each year are ranked according to their average assignment grade. Lastly, Table (4) shows the sex ratios of the classes, defined as the number of female students divided by the total number of students, where classes within each year are ranked according to their average assignment grade.

Tracking into HA classes can be used to identify the effect of peers because the rule induces a discontinuity in the relationship between assignment grade and class average grade at the assignment threshold. Since the discontinuity is the source of identifying information, some of the analysis that follows is restricted to students with assignment grades in a range close to the discontinuity point. Table (5) shows descriptive statistics for one such discontinuity sample, defined to include only students whose transformed assignment grade, $a_{itc} - s_t$, are in the interval [-0.5 , 0.5]. Slightly more than half of the students have assignment grades in this range. Table (6) shows descriptive statistics for the full sample. Comparison of the tables shows that the average characteristics of the classes in the discontinuity sample, except for grades, are remarkably similar to those for the full sample.

4 Empirical approach

My empirical approach exploits that there are 3-4 HA classes each year and the main determinant of which type of a class students are assigned to is the assignment grade and hence there is a discontinuity in the probability of being assigned to a HA class at the 60^{th} or 70^{th} percentile of the assignment grade. This cutoff in the sorting of students into HA classes constitutes a valuable source of identifying information. I exploit this to estimate a causal effect of classroom peers. To the best of my knowledge, this has not been done before. Students to the left of the assignment-determining threshold should provide valid counterfactual outcomes for students on the right side of the cutoff who were assigned to HA classes since the treatment status is randomized in a neighborhood of the threshold. I can therefore estimate the effect of class peers on academic outcomes by comparing outcomes for the students whose grades are just below and just above the threshold of getting into a HA class since they on average will have similar characteristics except for the treatment.

Since I am applying the fuzzy RD design, the probability of being assigned to a HA class is given by

$$E[H_{itc}|A_{itc}] = Pr[H_{itc} = 1|A_{itc} = a_{itc}] = \gamma + \delta \cdot 1\left(\frac{a_{itc} - s_t}{\sigma_{at}} \geq 0\right) + g\left(\frac{a_{itc} - s_t}{\sigma_{at}}\right), \quad (1)$$

where $1(\cdot)$ is the indicator function, taking the value one if the logical condition within the brackets holds and zero otherwise. H_{itc} is a treatment dummy taking the value one if student i in year t and class c was assigned to a HA class and zero otherwise, A_{itc} is the assignment grade of student i in year t and class c and σ_{at} is the standard deviation of the assignment grade at time t. $1\{s_t \geq c\}$ takes the value 1 if the assignment variable, the assignment grade a, exceeds the threshold, s_t, of having a higher probability of getting into a HA class which is given by the 60^{th} or 70^{th} percentile. $g(\cdot)$ is a control function, i.e. some low order polynomial in normalized assignment grade, $\frac{a_{itc}-s_t}{\sigma_{at}}$.

Assignment to HA classes can be represented by the following equation

$$H_{itc} = Pr[H_{itc} = 1|A_{itc} = a_{itc}] + u_{itc}$$

where u is an unobserved component which captures everything else influencing the class assignment decision, and academic achievement of students can be represented by the following equation

$$Y_{itc} = \alpha + \gamma_t + \beta X_c + \tau H_{itc} + f\left(\frac{a_{itc} - s_t}{\sigma_{at}}\right) + \epsilon_{itc}, \quad (2)$$

where Y_{itc} is an outcome variable for individual i in year t and class c, γ_t is a year specific effect, X_c is a vector of class characteristics and the effect of assignment grade is captured by the function $f\left(\frac{a_{itc}-s_t}{\sigma_{at}}\right)$, i.e. it is supposed to be an adequate description of $E[Y_{0itc}|A_i]$.

The key identification assumption that underlies the RD approach is that $f(\cdot)$ is a continuous function. Intuitively, the continuity assumption requires that differential assignment into classes is the only source of discontinuity in outcomes around the assignment threshold, 0, so that unobservables vary smoothly as a function of assignment grade and, in particular, do not jump at the cutoff. Formally, the conditional mean functions, $E\left[Y_{1i}|\frac{a_{itc}-s_t}{\sigma_{at}}\right]$ and $E\left[Y_{0i}|\frac{a_{itc}-s_t}{\sigma_{at}}\right]$, are continuous in $\frac{a_{itc}-s_t}{\sigma_{at}}$ at 0, or equivalently $E\left[\epsilon_i|\frac{a_{itc}-s_t}{\sigma_{at}}\right]$ are continuous in $\frac{a_{itc}-s_t}{\sigma_{at}}$ at 0. Under this assumption the treatment effect, τ, is obtained by estimating the discontinuity in the empirical regression function at the point where the treatment dummy, T, switches from 0 to 1 at the assignment threshold and can be given a causal interpretation.

H will be instrumented with the cutoff indicator C, which is defined as

$$C_{itc} = \begin{cases} 0 & \text{if} \quad a_{itc} < s_t \\ 1 & \text{if} \quad a_{itc} \geq s_t \end{cases}$$

since it captures the higher probability of being in a HA class at the assignment threshold, the 60^{th} or 70^{th} percentile of the assignment grade. The interpretation of equation (2) is that it describes the average potential outcomes of students under alternative assignments into HA classes, controlling for any other relationship between assignment grade and academic achievement. Since class types are not randomly assigned, it is likely to be correlated with the error component. OLS estimates of (2) will therefore not have any causal interpretation. The evaluation problem consists of estimating the effect of the assignment to a HA class on the outcome variable, i.e., τ .

The key identification assumption that underlies the RD approach is that $f(\cdot)$ is a continuous function. Intuitively, the continuity assumption requires that differential assignment into classes is the only source of discontinuity in outcomes around the assignment threshold, 0, so that unobservables vary smoothly as a function of assignment grade and, in particular, do not jump at the cutoff. Formally, the conditional mean functions, $E\left[Y_{1i}|\frac{a_{itc}-s_t}{\sigma_{at}}\right]$ and $E\left[Y_{0i}|\frac{a_{itc}-s_t}{\sigma_{at}}\right]$, are continuous in $\frac{a_{itc}-s_t}{\sigma_{at}}$ at 0, or equivalently $E\left[\epsilon_i|\frac{a_{itc}-s_t}{\sigma_{at}}\right]$ are continuous in $\frac{a_{itc}-s_t}{\sigma_{at}}$ at 0. Under this assumption the treatment effect, τ, is obtained by estimating the discontinuity in the empirical

regression function at the point where the treatment dummy, T, switches from 0 to 1 at the assignment threshold and can be given a causal interpretation.

As shown in Lee (2008) and Lee and Lemieux (2009), smoothness of the density of the treatment-determining variable is sufficient for the continuity assumption to hold. In my case, this assumption explicitly allows for students to have some control over their value of the assignment grade. As long as this control is imprecise, assignment to HA classes will be randomized around the threshold. In my case, the continuity of the assignment grade density function also directly ensures that assignment into HA classes is randomized close to the assignment threshold. An additional concern would be imperfect compliance with the treatment rule, but in my study there is not much scope for this. Each classroom has limited space which is in most cases fully utilized and there is no possibility to switch classes if there is not enough space in another class. Also, a student's outside options are scarce if she decides to leave the school because she was not assigned to a HA class. The student will need to wait at least one semester to get into another school and for a whole year if she wants to get into the most popular ones.

It is also helpful to consider how reasonable the continuity assumption is in the context of this paper? This system of streaming students into HA and normal classes has never been official and students were therefore, in most cases, not aware of the system until they had started their studies. If students were aware of the system's existence prior to enrolment they obviously had an incentive to affect the way school administrators assigned them into classes, and presumably also some control over this. However, it seems implausible that this control was perfect, so the key identifying assumption is likely to hold here. Furthermore, assignment grade is determined after students receive their grades on the standardized exams and school exams so they were unlikely to know the exact location of the HA class cutoff even if they wanted to make sure that they managed to reach the cutoff.

One might also worry that school administrators had incentives to alter the cutoffs to benefit students they favored. It is unlikely, however, that this kind of manipulation would have occurred. For instance, in order for administrators to have used the cutoffs to benefit particular students they favored, there would have had to be places on the support of the student grade distribution where favored students had a systematically higher density than other students.

A final potential concern is that other school policies are also related to the same grade cutoffs. To my knowledge, however, there are no programs that use the same cutoff.

Peer effects provide therefore an example of how fuzzy RD can be analyzed in an instrumental variable framework where the IV estimates can be given causal interpretation. In this case, IV estimates of equation (2) use discontinuities in the relationship between assignment grade and assignment into HA classes to identify the causal effect of peers ability at the same time that any other relationship between assignment grade and academic achievement measured by the end of the first year is controlled for by including a smooth function of assignment as a control. In practice, this includes linear, polynomial and local linear functions of assignment grade.

Because there are relatively few observations in a local neighborhood of the assignment threshold, the control function approach is my preferred method in my RD analysis. The disadvantage of this approach is that it becomes a major concern whether the specification of the control function, $f(\cdot)$, which determines the slope and the curvature of the regression line and affects therefore the estimated treatment effect, is correct. I therefore use a couple of different specifications when using an extended support. As a further specification test, I will also estimate the effect of being in a HA class using only observations that are +/– 5 percentage points from the assignment grade threshold without any control functions for assignment grade. The idea behind the RD design is that this discontinuity sample will be a close approximation to a randomized trial and therefore it is unnecessary to include the control function. Consequently, the estimate from the discontinuity sample should now be equal (apart from sampling variability) to the estimate from the control function approach, unless the control function is misspecified. However, since the slope of the relationship between the assignment grade and academic achievement is rather steep around the discontinuity, the discontinuity sample would need to be very small for this estimate to give an accurate description of the causal effect of being assigned to a HA class. I therefore also use local linear regression in samples around the discontinuity (+/– 5 percentage points from the assignment grade threshold), which amounts to running simple linear regressions allowing for different slopes of the regression function in the neighborhood of the assignment-threshold. I follow the suggestions by Imbens and Lemieux (2008) and use a rectangular kernel, i.e. equal weights for all observations in the sample used. A linear control function should be able to capture any other relationship between assignment grade and academic achievement in such a close proximity to the threshold but I also show estimates where I include a second order polynomial in normalized assignment grade as a control.

5 Results

The first crucial assumption for being able to apply the RD design is that there is an observable assignment variable on which assignment is based and that there is a discontinuity at some cutoff value of the assignment variable in the level of treatment. In this case the assignment variable is the assignment grade and the threshold is the assignment grade at the 60^{th} or the 70^{th} percentile (depending on which year we consider). This assignment rule is graphically displayed in Figure (6) and fits the treatment allocation rule of the fuzzy RD design. The assignment as a function of the normalized assignment grade, $\frac{A-S}{\sigma_A}$, contains a jump at a known threshold value for $\frac{A-S}{\sigma_A}$, namely 0, so this first assumption is fulfilled. This also provides an informal way of sensing how large the jump in the probability of being assigned to a HA class, δ, is at the cutoff point, and what the functional form $g(\cdot)$, in equation (1), looks like.

As a first exploration for a possible effect of classroom peers on educational outcomes, I plot the average grades by the end of the first year as a function of the average assignment grade and see whether they exhibit a similar trend around the threshold value. I do this by using binned local averages, i.e., the assignment grade is binned so that all grades between x and y were assigned the assignment grade of $\frac{x+y}{2}$, grades between y and z were assigned the assignment grade of $\frac{y+z}{2}$, and so forth. I do this for spring exam results and year grade in Figures (2) and (3), respectively.

The figures show that there is a discontinuity in normalized spring exam result and year grade around the assignment threshold and therefore present evidence that academic achievement, as measured by spring exam results or year grade, is affected by being assigned into a HA class. A second exploration for a possible effect of ability peer effects is to compare outcomes of discontinuity samples around the HA class threshold. Table (7) compares normalized spring exam results and normalized year grades for HA and normal classes. I use 5 discontinuity samples and there is considerable difference between the outcomes for the two class types in all of them. The smallest discontinuity sample, the $+/-0.5$ % sample, includes all observations that are in the range of $[-0.05, 0.05]$ % of the transformed assignment grade, $a_{itc} - s_t$, and this sample should therefore be a close approximation to a randomized trial. Although the figures and the table suggest that class type, induced by the percentile assignment rule, affects student achievement measured by the end of the first year, they do not provide a framework for formal statistical inference. Table (8) shows analytical results from instrumental variable regressions of academic achievement on class type (i.e., equation (2)). The control function approach is my preferred method since there is only a limited number of observations close to the threshold in the data set (i.e. there are only 663 observations within +/– 5 percentage points from the threshold).

Specifications of the control function include a first-order up to a third-order polynomial in normalized assignment grade (see columns 1-3) as a way of testing whether the estimate of the effect of being in a HA class is sensitive to the different specifications of the control function. I also include local linear estimates, using only observations that are +/– 5 percentage points from the assignment grade threshold with only linear control functions for assignment grade (see column 4). In addition I show estimates where I include a second order polynomial in normalized assignment grade as a control (see column 5). I also run regressions where the slope of the regression function differ on both sides of the cutoff point by including interaction terms between H and A. However, the difference between the polynomials when I include higher than first order polynomials turns out to be insignificant. It seems therefore that the functional form of the control function is the same on both sides of the cutoff if it is of higher order than one. I therefore only include estimates that have different linear slopes since I obtain more efficient estimates of the treatment effect by not including unnecessary constraints (see columns 11 and 12).

Looking at Table (8) we see that when using a parametric 2SLS setup in the full sample, instrumental variable estimates of the effect of being assigned to a HA class on spring exam results and year grade are positive and significant for all specifications of the control function. The point estimates obtained for spring exam results when using a second and third degree polynomial are very similar, indicating a positive effect of 0.246 and 0.224 standard deviations, respectively, and 0.213 when using a third degree polynomial and controls. The effects are statistically significant at the 1 percent level when using a second order polynomial and at the 5 percent level when using a third order polynomial. The same holds for year grades, the estimates obtained when using a second and third degree polynomial are again very similar, indicating a positive effect of 0.235 and 0.224 standard deviations, respectively and 0.221 when using a third degree polynomial and controls. These estimates are all significant at the 1 percent level.

For the discontinuity sample the standard errors are more than 50% larger than when using the full sample. This explains why the control function approach is my preferred method: it is much more efficient than just comparing the average outcomes in a small neighborhood on either side of the treatment threshold. The effect is still positive, but only significant (at the 10 percent level) for year grade. For the +/– 10 percentage window there is also a positive and statistically significant effect at the 10 percent level. The estimates for the +/– 2.5 percentage window are very imprecise and therefore not statistically distinguishable from zero. The fact that we obtain positive and significant estimates fits with the visual results which suggested that there was a positive treatment effect on academic performance from assigning students to HA classes.

The treatment effect is quite large, the difference between the normalized assignment grade of normal and HA classes is 0.55 but the treatment effect for spring exam results and year grade is approximately 0.213-0.246 and 0.221-0.235, respectively. This suggests that if a student with assignment grade just below the HA class threshold would instead of being assigned to a normal class go to a HA class, where students have assignment grades that are on average 0.55 standard deviations higher than in normal classes, this would lead to a more than 0.2 standard deviation increase in spring exam and year grade results. In other words, increasing academic ability of peers by one standard deviation increases one's own academic performance by approximately 0.42 standard deviations.

Internal validity of the RD approach is based on the local continuity assumption, i.e., that the conditional expectation of the outcome variable is continuous around the discontinuity point. It is therefore very important to obtain clear indication that this holds. This assumption cannot be tested in general but a number of validity and sensitivity tests have been developed to bolster the credibility of the RD estimates. First, it is helpful to consider why this assumption might break. Economic behavior can invalidate the assumption of local continuity, this can for instance come about in cases where individuals with a stake in T_i are able to manipulate the assignment variable in order to affect whether or not they fall on one side of the cutoff or the other. Also, if administrators can strategically choose which cutoff point to pick or what assignment variable to use, then comparability near the threshold may be violated. Both types of behavior could lead to sorting of individuals close to the threshold. Sorting of individuals around the cutoff may lead to different average characteristics of those above and below the threshold so the internal validity of the results would break in this case. Another reason for why the local continuity assumption might break is if there are other discontinuous programs using the same assignment variable and cutoff value. This would lead to changes at the cut-off value that may affect the outcome, and these effects may be attributed erroneously to the treatment of interest.

A first validity check for the local continuity assumption is to look at the density of the assignment variable close to the threshold, X_0. This is of importance since manipulation of the assignment grades on which the HA class assignment was based on would cast doubts on the internal validity of the research design. A discontinuous jump in the density of the assignment variable at the discontinuity would be considered as suggestive of sorting behavior and hence a violation of the RD assumptions. In order to show that this is not the case I present a histogram of the assignment variable in Figure (4). Visual inspection does not show any unusual jump at the threshold, suggesting that sorting should not be a problem in this case.

Another check to test for imbalance of relevant characteristics and hence the validity of the local continuity assumption is to compare average characteristics of individuals on either side of the threshold to make sure that they are observationally similar. This is done in Table (9) where I compare sex ratios, average age, class size, the ratio of students living in the capital region and the ratio of students from a school in the capital region for the two groups. There are small differences in sex ratios and class sizes but no difference for the other measures. Therefore I included sex ratios and class size as class characteristics (X_c) when estimating equation (2). Even if we do not find any difference in observational characteristics there could be discontinuities in unobservable characteristics around the cutoff and if this is the case and the unobserved characteristic is related to the outcome variable the RD assumptions do not hold. This can be tested by applying a test for imbalance in relevant variables (van der Klaauw, 2002, 2008) where I also run the regressions including observed characteristics as controls. The only thing to gain from controlling for observable characteristics when estimating treatment effects in a RD design, given that they have explanatory power, is reduction in sampling variability. If the RD estimates are sensitive to inclusion of observed characteristics as controls, this would be taken as suggestive of violation of the continuity assumptions. I therefore include the ratio of students from elementary schools in the capital region and ratio of students living in the capital region. The estimates hardly change when we add these covariates but the standard errors are smaller as expected. In other words, this suggests that assignment into HA classes is in fact “as good as” random around the threshold.

There are also certain conditions that can make sorting less likely. If the assignment rule is not known and hard to uncover, if the location of the cutoff is unknown or uncertain, if the assignment variable cannot be manipulated or if there is insufficient time for agents to do so it is less likely that we have a sorting problem. In this study the assignment rule specifies that the cutoff is the assignment grade at the 60^{th} or 70^{th} percentile so it is difficult for applicants to know where the cutoff lies and there is limited scope for strategic behavior by the administrators since they do not have anything to say about which assignment variable to use or where the cutoff lies. This suggests that it is unlikely that sorting would cause problems in this study.

6 Conclusions

In this paper, I have estimated ability peer effects using data for five cohorts of age 16 in an Icelandic high-school where I measure peers' ability by their academic ability as recorded by standardized test scores and test scores from their previous schools (elementary school). The outcome variable of interest is their academic performance by the end of their first year of study, measured by spring exam results and year grade, which is an overall measure of how they have done on homework assignment, quizzes, Christmas exams etc. during the year.

From a methodological perspective, I view my main contribution to be the approach taken to measure peer effects, where student assignment into HA classes constitutes the source of identifying information. As far as I know, this has never been done before.

In terms of findings, my results suggest that assigning students to classes with peers of higher academic ability increases their own academic performance. The conjecture that peers' ability cause differentials in academic performance is therefore substantiated empirically; tracking students into classes does seem to exacerbate inequality among students who ex ante are of equal ability. In more detail, my estimates suggest that a 1 standard deviation increase in the average ability of peers would increase one's own outcomes by approximately 0.42 standard deviations. The previous literature finds peer effects that range from close to zero (Sanbonmatsu et al., 2004) to about 0.5 standard deviations for a one standard deviation change in the peer measure (Hoxby 2000; Boozer and Cacciola 2001). My results therefore fall within this range but are close to the upper end. This is consistent with the fact that the estimated peer effects were somewhat greater in those previous studies where it was possible to identify classmates.

7 References

Ammermueller, Andreas and Jrn-Steffen Pischke. (2009). “Peer effects in European primary schools: Evidence from PIRLS. *Journal of Labor Economics* 27(3), 315-348.

Bui, Sa A, Steven G. Craig and Scott A. Imberman. (2011). “Is gifted education a bright idea? Assessing the impact of gifted and talented programs on achievement. NBER Working Paper No. 17089.

Boozer, Michael A. and Stephen E. Cacciola (2001). “Inside the 'Black Box' of Project STAR: Estimation of peer effects using experimental data. Economic Growth Center Center Discussion Paper No. 832, Yale University.

Duflo, Ester, Pascaline Dupas, and Michael Kremer. (2008). “Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. NBER Working Paper No. 14475.

Gibbons, Stephen and Shqiponja Telhaj. (2008). “Peers and Achievement in England's Secondary Schools. SERC Discussion Paper.

Goux, Dominique and Eric Maurin. (2007). “Close Neighbors Matter: neighbourhood effects on early performance at school. *The Economic Journal* 117, 1193-1215.

Graham, Bryan. (2008). “Identifying social interactions through conditional variance restrictions. *Econometrica* 76 (3): 643-660.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. (2003). “Does peer ability affect student achievement? *Journal of Applied Econometrics* 18, 527-44.

Hoxby, C. (2000). “Peer Effects in the Classroom: Learning from Gender and Race Variation. NBER Working Paper 7867.

Hoxby, Caroline and G. Weingarth. (2006). “Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects. mimeo.

Imbens, Guido W., and Thomas Lemieux. (2008). "Regression discontinuity designs: A guide to practice *Journal of Econometrics* 142(2): 615-635.

Lavy, Victor, Olmo, S., and Weinhardt, F. (2009). "The good, the bad and the average: evidence on the scale and nature of ability peer effects in schools. NBER Working Paper No. 15600.

van der Klaauw, Wilbert. (2002). "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach. *International Economic Review* 43(4), 1249-1287.

van der Klaauw, Wilbert. (2008). "Breaking the Link between Poverty and Low Student Achievement: An Evaluation of Title I. *Journal of Econometrics* 142(2), 731-756.

Lavy, Victor, Olmo Silva and Felix Weinhardt. (2009). "The good, the bad and the average: evidence on the scale and nature of ability peer effects in schools. forthcoming in the Journal of Labor Economics.

Lee, David S. (2008). "Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics* 142(2), 675-697.

Lee, David S. and Thomas Lemieux. (2009). "Regression Discontinuity Designs in Economics. NBER Working Paper 14723.

Manski, Charles F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* 60(3), 531-542.

Marsh, Herbert W. (2005). "Big Fish Little Pond Effect on Academic Self-concept: Cross-cultural and Cross-Disciplinary Generalizability. Paper presented at the AARE Conference.

McEwan, Patrick J. (2003). "Peer effects on student achievement: Evidence from Chile. *Economics of Education Review* 22, 131-41.

Sacerdote, Bruce. (2001). "Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics* 116(2): 681-704.

Sanbonmatsu, Lisa, Jeffrey R. Kling, Greg J. Duncan and Jeanne Brooks-Gunn. (2006). "Neighbourhoods and Academic Achievement: Results from the Moving to Opportunity Experiment. *Journal of Human Resouces* 41(4), 649-691.

Schindler Rangvid, Beatrice. (2007). "School composition effects in Denmark: Quantile regression evidence from PISA 2000. *Empirical Economics* 33, 359-88.

Schneeweis, Nicole and Rudolf Winter-Ebmer. (2007). "Peer Effects in Austrian Schools. *Empirical Economics* 32(2-3), 387-409.

Zimmerman, David J. (2003). "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics* 85(1), 9-23.

Table 1: Assignment grade for each class

	Year				
Class	1995	1996	1997	1998	1999
1	8.96*	8.79*	8.72*	8.86*	8.85*
2	8.8*	8.74*	8.7*	8.82*	8.84*
3	8.79*	8.73*	8.65*	8.8*	8.79*
4	7.79	7.57	8.11	8.73*	8.78*
5	7.77	7.56	8.06	8.03	8.18
6	7.74	7.48	7.92	7.97	8.15
7	7.74	7.48	7.84	7.96	7.95
8	7.65	7.47	7.83	7.93	7.94
9	7.61	7.45	7.82	7.92	7.91
10	7.55	7.37	7.77	7.91	7.88
Total	8.05	7.88	8.17	8.29	8.34

* High-ability classes

Assignment grade is defined as the average of student's results in Mathematics, Icelandic, English and Danish on the standardized tests for students in the 10^{th} grade and school grades in these same subjects.

Table 2: Normalized assignment grade for each class

Class	Year 1995	1996	1997	1998	1999
1	0.34*	0.33*	0.21*	0.47*	0.24*
2	0.22*	0.29*	0.19*	0.43*	0.23*
3	0.21*	0.29*	0.14*	0.41*	0.21*
4	-0.52	-0.64	-0.38	0.32*	0.2*
5	-0.54	-0.65	-0.42	-0.45	-0.15
6	-0.56	-0.71	-0.55	-0.52	-0.17
7	-0.57	-0.72	-0.64	-0.53	-0.29
8	-0.63	-0.72	-0.65	-0.56	-0.29
9	-0.66	-0.74	-0.66	-0.57	-0.31
10	-0.7	-0.81	-0.71	-0.59	-0.33
Total	-0.33	-0.39	-0.32	-0.16	-0.06

* High-ability classes Normalized assignment grade is defined as assignment grade minus the high-ability assignment threshold where the probability of being assigned to a high-ability class jumps and divided by the standard deviation of assignment grade, i.e., $\frac{a_{itc}-s_t}{\sigma_{at}}$.

Table 3: Number of students in each class

Class	Year 1995	1996	1997	1998	1999
1	28*	26*	28*	28*	28*
2	27*	28*	28*	28*	28*
3	28*	28*	28*	28*	28*
4	27	26	27	28*	28*
5	25	28	27	28	28
6	27	27	25	28	28
7	26	27	25	27	27
8	27	24	25	27	28
9	27	24	23	28	28
10	27	26	24	29	28
Total	269	264	260	279	279

* High-ability classes. Classes within each year are ranked according to the average assignment grade.

Table 4: Sex ratio in each class

	Year				
Class	1995	1996	1997	1998	1999
1	0.46*	0.54*	0.50*	0.46*	0.46*
2	0.44*	0.57*	0.50*	0.54*	0.71*
3	0.46*	0.46*	0.46*	0.5*	0.46*
4	0.48	0.31	0.56	0.68*	0.54*
5	0.36	0.32	0.52	0.64	0.68
6	0.37	0.3	0.52	0.36	0.64
7	0.46	0.33	0.52	0.44	0.52
8	0.44	0.25	0.56	0.52	0.50
9	0.41	0.33	0.43	0.54	0.50
10	0.44	0.27	0.46	0.45	0.54
Total	0.43	0.37	0.50	0.51	0.56

* High-ability classes.

Sex ratio is defined as the number of female students divided by the total number of students.

Classes within each year are ranked according to the average assignment grade.

Table 5: Descriptive Statistics
+/– 5 Discontinuity Sample

	Mean	Standard Deviation
Ratio living in the capital region	0.951	0.0472
Ratio from a school in the capital region	0.907	0.0553
Sex ratio	0.494	0.0982
Class size	27.341	1.5018
Age	15.986	0.1481
Standardized exam in Icelandic	8.066	0.6356
Standardized exam in Math	8.206	0.9097
Standardized exam in Danish	8.434	0.8057
Standardized exam in English	8.379	0.7749
School grade in Icelandic	8.377	0.6377
School grade in Math	8.613	0.7341
School grade in Danish	8.588	0.7101
School grade in English	8.685	0.6307
Assignment grade	8.428	0.3064
Normalized assignment grade	-0.027	0.2543
Year grade	7.478	1.2566
Normalized year grade	0.280	1.0328
Spring exam result	6.955	1.3025
Normalized spring exam result	0.223	0.7901
Class type (high-ability = 1)	0.467	0.4992
Over threshold	0.504	0.4992

Notes: Number of observations in the discontinuity sample is 723. Sex ratio is defined as the number of female students divided by the total number og students.

Table 6: Descriptive Statistics
Full Sample

	Mean	Standard Deviation
Ratio living in the capital region	0.954	0.0460
Ratio from a school in the capital region	0.908	0.0545
Sex ratio	0.476	0.1021
Class size	27.089	1.4899
Age	16.009	0.2140
Standardized exam in Icelandic	7.659	1.3931
Standardized exam in Math	7.747	1.6116
Standardized exam in Danish	7.962	1.5553
Standardized exam in English	8.077	1.4468
School grade in Icelandic	7.977	1.4047
School grade in Math	8.133	1.5312
School grade in Danish	8.079	1.5445
School grade in English	8.340	1.4175
Assignment grade	8.007	1.2915
Normalized assignment grade	-0.364	1.0043
Year grade	7.112	1.6338
Normalized year grade	0	1.3409
Spring exam result	6.556	1.6852
Normalized spring exam result	0	0.9985
Class type (high-ability = 1)	0.340	0.4770
Over threshold	0.365	0.4816

Notes: Number of observations in the full sample is 1352. Sex ratio is defined as the number of female students divided by the total number og students.

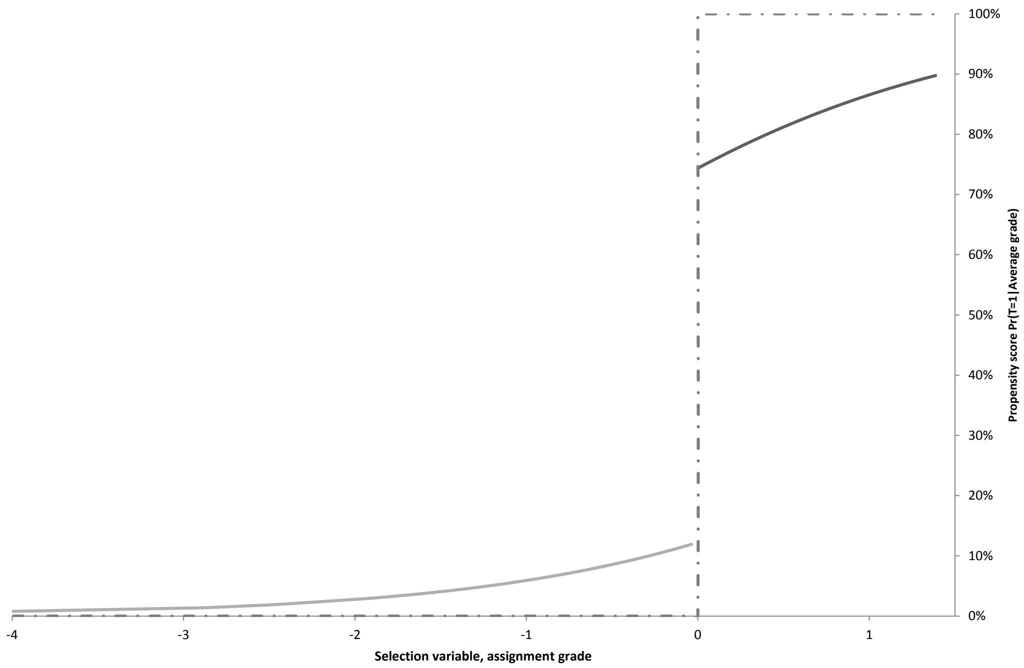

Figure 1: Probability of being assigned to a high-ability class in 1995-1999 as a function of normalized assignment grade

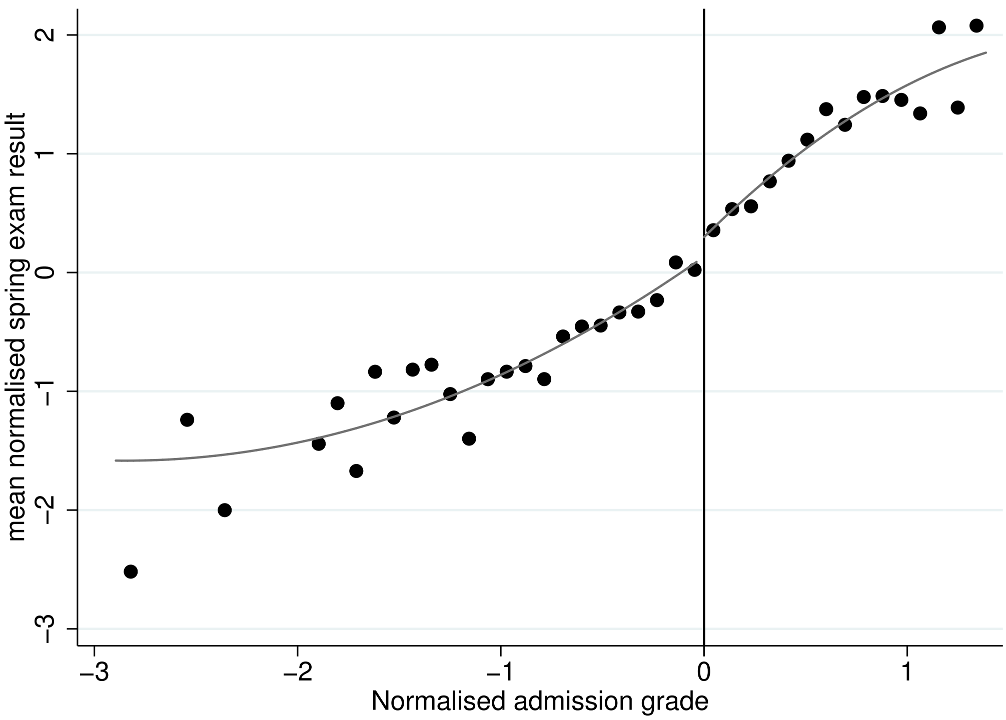

Figure 2: Spring exam results as a function of assignment grade in 1995-1999, using binned local averages

Assignment grade is binned so that all grades between x and y were assigned the assignment grade of $\frac{x+y}{2}$, grades between y and z were assigned the assignment grade of $\frac{y+z}{2}$, and so forth.

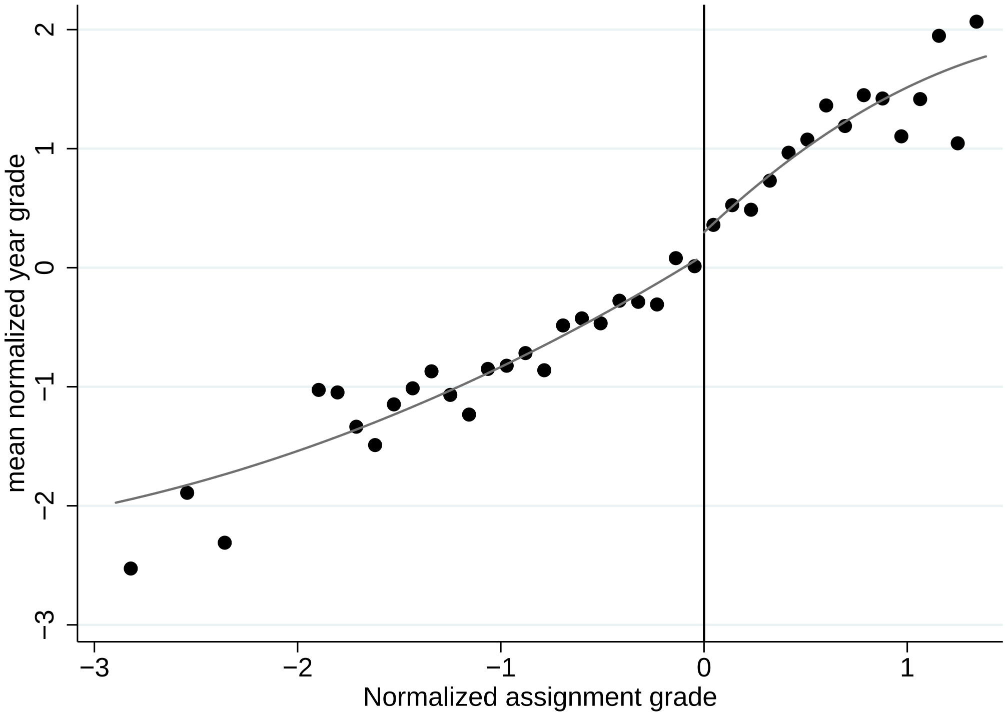

Figure 3: Year grade results as a function of assignment grade in 1995-1999, using binned local averages

Assignment grade is binned so that all grades between x and y were assigned the assignment grade of $\frac{x+y}{2}$, grades between y and z were assigned the assignment grade of $\frac{y+z}{2}$, and so forth.

Table 7: Comparison of outcomes for discontinuity samples in HA and normal classes

	Normal classes		HA classes	
	Spring exam results	Year grades	Spring exam results	Year grades
Discontinuity samples:				
$+/-10$ %	-0.22	-0.19	0.61	0.54
$+/-5$ %	0.00	0.02	0.47	0.43
$+/-2.5$ %	0.07	0.05	0.41	0.37
$+/-1$ %	0.03	0.06	0.42	0.36
$+/-0.5$ %	0.20	0.24	0.39	0.33

Note: The full sample includes 1352 observations. The $+/-10$ % sample includes all observations that are in the range of $[-1.0, 1.0]$ of the transformed assignment grade, $a_{itc} - s_t$, and there are 1158 such observations. The $+/-5$ % sample includes all observations that are in the range of $[-0.5, 0.5]$ of the transformed assignment grade and there are 723 such observations. The $+/-2.5$ % sample includes all observations that are in the range of $[-0.25, 0.25]$ of the transformed assignment grade and there are 431 such observations. The $+/-1$ sample includes all observations that are in the range of $[-0.01, 0.01]$ of the transformed assignment grade and there are 146 such observations. The $+/-0.5$ % sample includes all observations that are in the range of $[-0.05, 0.05]$ % of the transformed assignment grade and there are 77 such observations.

Table 8: Instrumental variables estimates

	1	2	3	4	5	6	7	8	9	10	11	12
Spring exam result	0.367***	0.246***	0.224**	0.306	0.290	0.213**	0.290	0.273	0.185*	0.549	0.327***	0.279
	(0.0936)	(0.0930)	(0.1033)	(0.1897)	(0.1897)	(0.0988)	(0.1818)	(0.1811)	(0.995)	(0.6245)	(0.0891)	(0.1969)
Year grade	0.305***	0.235***	0.224***	0.312*	0.312*	0.221***	0.303*	0.304*	0.184*	0.084	0.276***	0.312
	(0.0773)	(0.0796)	(0.0868)	(0.1706)	(0.1777)	(0.0856)	(0.1695)	(0.1769)	(0.1845)	(0.5569)	(0.0764)	(0.1879)
Sample	Full	Full	Full	+/– 5	+/– 5	Full	+/– 5	+/– 5	+/– 10	+/– 2.5	Full	+/– 5
Tranformed assignment grade polynomial	First	Second	Third	First	Second	Third	First	Second	Second	Second	First	First
Controls	No	No	No	No	No	Yes	Yes	Yes	No	No	No	No

Notes: Standard errors are clustered at the class level and are within parentheses. Each entry is separate regression. The full sample includes 1290 observations. The $+/-5$ sample includes all observations that are in the range of $[-0.5, 0.5]$ of the transformed assignment grade, $a_{itc} - s_t$, and there are 712 such observations. The $+/-10$ sample includes all observations that are in the range of $[-1, 1]$ of the transformed assignment grade and there are 1136 such observations. The $+/-2.5$ sample includes all observations that are in the range of $[-0.25, 0.25]$ of the transformed assignment grade and there are 422 such observations.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table 9: Average characteristics of students in normal and high-ability classes

	Class type	
	Normal	High-ability
Sex ratio	0.46	0.51
	(0.11)	(0.07)
Average age	16.03	15.97
	(0.24)	(0.16)
Class size	26.72	27.83
	(1.44)	(0.5)
Ratio living in	0.96	0.95
the capital region	(0.20)	(0.22)
Ratio from a school	0.91	0.90
in the capital region	(0.28)	(0.30)
Assignment	7.61	8.73
grade	(1.33)	(0.80)
Normalized assignment	-0.68	0.23
grade	(0.98)	(0.75)
Spring exam	6.01	7.59
result	(1.57)	(1.36)
Normalized spring exam	-0.33	0.62
result	(0.93)	(0.81)
Year grade	6.65	7.99
	(1.59)	(1.30)
Normalized	-0.29	0.54
year grade	(0.97)	(0.80)

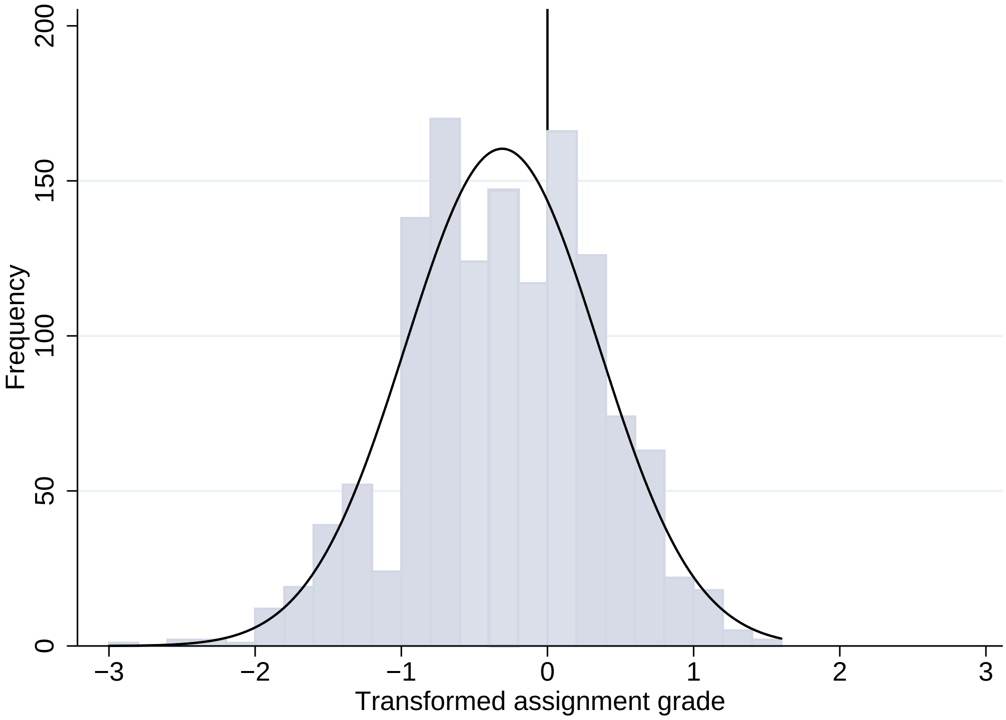

Figure 4: Histogram of transformed assignment grade, $a_{itc} - s_t$, in 1995-1999. The bin-width in the histogram is 0.2 and no bin counts observations from both sides of the cutoff.