The Impact of School Tracking and Peer Quality on Student Achievement: Regression Discontinuity Evidence from Thailand

Meradee Tangvatcharapong *

November 2020

Abstract

A common education practice around the world is to track students into classrooms based on ability. This paper estimates the impact of being tracked into classrooms with high-ability peers using administrative data from public middle schools in Thailand, where students are tracked into classrooms based on a preliminary exam taken before the seventh grade. Importantly, all teachers, curriculum, and textbooks are identical throughout classrooms. To identify effects, I apply a regression discontinuity design that compares the academic outcomes of students just above and below the classroom threshold. Results indicate that significant increases in peer quality do not improve student GPA.

JEL Codes: I21, I28, J21, J24

*Meradee Tangvatcharapong is a PhD Candidate at Texas A&M University. Email: meradee_tang@tamu.edu

1 Introduction

Across the world, a common educational practice is to track students into different classrooms based on ability. For example, the OECD reports that 95 percent of students in the United Kingdom, Ireland, New Zealand, Australia, Israel, Albania, Kazakhstan, Singapore, Russia, and Malaysia attended schools where students were grouped by ability across classrooms (OECD, 2013). While student tracking is generally less common in the U.S., it is still widely used in some areas. Per the National Assessment of Education Process, 75 percent of U.S. schools track students by ability for 8th-grade mathematics (Loveless, 2013). However, despite the popularity of tracking, there is still debate over the merits of tracking students by ability. Of particular concern is whether tracking could harm students who were tracked into lower ability classrooms through increased exposure to lower-ability peers. Unfortunately, there is little direct evidence on whether these differences in peer exposure induced by tracking cause subsequent differences in student achievement. This is largely due to the fact that many programs that assign higher-ability students to different classrooms, such as gifted and talented programs, also often expose those students to different classroom settings, such as more intensive curriculum and higher-quality teachers. This makes it difficult to identify whether and how much the most common change induced by tracking, namely changes in peer quality, affects student achievement. The purpose of this paper is to estimate the effects of being tracked into higher-ability classrooms in a setting where only peer quality changes, thereby separating the effects of peers from other confounding factors. In doing so, I also speak to the underlying reasons for the mixed evidence in the literature.

To this end, I apply a regression discontinuity design using administrative data from public middle schools in Thailand where public schools regularly sort students into classrooms based on ability. To measure student ability, Thai public middle schools usually have students sit for a preliminary exam before the start of seventh grade. School officials then use the score from the preliminary exam as a proxy for student ability and sort students into classrooms based on this preliminary exam score. This allows me to exploit the resulting cutoffs between classrooms to employ a regression discontinuity approach that compares the academic outcomes of students just above and below the cutoffs to identify the effects of peer quality in the tracking system. There are several reasons why the institutional setting here is ideal for a regression discontinuity design. First, these cutoffs between classrooms are not known to the students until after the exam is taken and classrooms are assigned, making it difficult if not impossible to manipulate one's position relative to a classroom cutoff. Second, classrooms above and below the cutoffs are required to follow the same curriculum, take the same exam, and have nearly identical sets of teachers. As a result, this setting supports the identifying assumption that all determinants of achievement other than peer quality vary smoothly across the classroom cutoff. I provide empirical evidence supporting this assumption in Section 4.

For the analysis, I implement the aforementioned design using administrative student data from four public middle schools in Thailand. My data set contains the preliminary score, classroom assignment, GPA, classroom timetable, teacher assignment, and characteristics of 1,602 seventh-grade students. The main outcome of interest is the seventh-grade cumulative GPA. Importantly, GPA in Thailand is based primarily on student performance in multiple-choice exams for which there is no grade curving. Students in the same school also take the same exams regardless of their assigned classroom and teachers. As a result, there is very little scope for teacher bias or subjectivity to affect GPA in this context.

Consistent with the classroom allocation mechanism, I first show that scoring just above the cutoff does increase the likelihood of being assigned to the higher-ability classroom by 80 percentage points. However, results indicate there is no discontinuity in student performance at the cutoff. This indicates that being tracked into classrooms with significantly higher-ability peers does not lead to higher achievement. Specifically, my 2SLS estimates show that assignment to a higher-ability classroom is associated with a 0.94 standard deviation increase in peer quality, as measured by performance on the seventh-grade preliminary exam. However, this exposure is associated with a statistically insignificant 0.08 standard deviation reduction in performance, as measured by GPA. Importantly, this result is robust to the bandwidth size as well as the inclusion of student characteristics. Additionally, the result is also robust to the inclusion of teacher fixed effects. This is consistent with the fact that students have the same or similar teachers across classroom cutoffs. as well as the fact that Thai teachers in my data set do not systematically choose classrooms, i.e. higher-quality teachers do not only teach classrooms just above the cutoffs.

In addressing the effects of being tracked into classrooms with higher-ability peers, the paper is most closely related to a paper by Vardardottir (2013). That paper uses student data from Iceland and also aims to identify the impact of being tracked into classrooms with higher-ability peers. The major difference between this paper and that one is that I observe a clear, visually compelling discontinuity in the likelihood of being placed in a classroom with higher-achieving peers at the cutoff. In contrast, there is no such discontinuity in the data underlying Vardardottir (2013).[1] As a result, a major contribution of my paper is to identify the effects of being tracked into a higher-ability classroom using a clean regression discontinuity framework. This enables me to give estimates a causal interpretation under a reasonable identifying assumption. In addition, this paper is the first to identify the effect of peer quality shifts due to tracking in Asia, where tracking is very common.

In addition to providing estimates in a clean regression discontinuity framework, another advantage of my study is that I am able to rule out positive effects of modest size. For example, Duflo, Dupas, and Kremer (2011) performed a field experiment in Kenya that enabled a regression discontinuity study of the effects of being tracked into higher-ability classrooms on student achievement. Similar to this study, they reported no statistically significant effects from an increase of one standard deviation in peer quality and ruled out effects larger than 0.21 standard deviations. By comparison, estimates in this study enable me to rule out effects of only 0.08 standard deviations.

This study is also directly related to the literature on the general effectiveness of tracked classrooms and gifted and talented programs (Bui, Craig, and Imberman, 2014; Card and Giuliano, 2016; Booij, Haan, and Plug, 2017; Cohodes, 2020). The results in this paper support

[1]This is evident in the local averages shown in Figure 1 on page 115

the finding in Bui, Craig, and Imberman (2014) which used a regression discontinuity design to estimate the impact of gifted and talented programs from a large school district in the US and found that achievement does not improve for students placed in gifted and talented programs. The results here also speak to the finding in Card and Giuliano (2016) which used a regression discontinuity design to study the impact of gifted classrooms in a large school district in the US. They found that the impact of the gifted classrooms was minimal for white students but large for minority students. Since white and minority students experience the same curriculum, teachers, and peers, they concluded that the large effects on minority students were likely from the removal of low teacher expectations and negative peer pressure. As Thailand is a relatively homogeneous country, it is not surprising that the result in this paper is in line with the results of white students in Card and Giuliano (2016). Additionally, estimates in this study also enable me to rule out the effects of the magnitude found by some studies on gifted and talented programs. For example, Booij, Haan, and Plug (2017) reports that a gifted and talented program in the Netherlands increased student GPA by 0.2 standard deviations. My result here suggests that a large portion of the positive effects found in Booij, Haan, and Plug (2017) are likely due to specific features of the gifted and talented program, such as the curriculum and teacher quality, rather than the change in peer quality.

As attending higher-quality schools is often associated with an increase in peer quality, this paper also speaks to the literature on returns to school quality. The results in this paper again suggest that the positive effects found by some studies (Shi, 2019; Pop-Eleches and Urquiola, 2013; Park, Shi, Hsieh, and An, 2009) could perhaps be attributed more to features of the schools other than higher peer quality in the classroom. It also provides a plausible explanation as to why many papers (Allensworth, Moore, Sartain, and de la Torre, 2017; Dee and Lan, 2015; Lucas and Mbiti, 2014; Dobbie and Fryer Jr, 2014; Abdulkadiroğlu, Angrist, and Pathak, 2014; Clark, 2010) have found selective schools to have little effect on student achievement. Indeed, the findings of this paper would predict that same result, unless selective schools also offered better teachers or other input into education production. In this way, this paper also complements the finding in

Hoekstra, Mouganie, and Wang (2018) and Jackson (2013) that the returns to high school quality are likely the result of features of the schools other than peer quality. Additionally, this paper also complements the large literature on peer effects more generally.[2]

The results of this paper also have important implications for both parents and policymakers. Recent evidence suggests that in choosing schools, parents put much weight on peer quality (Abdulkadiroglu, Pathak, Schellenberg, and Walters, 2017). However, results shown in this paper suggest that parents would be better off making decisions on factors other than peer quality, such as teacher quality or curriculum. Similarly, they also suggest that educators and policymakers should put more emphasis on other factors believed to improve student performance and less emphasis on the role of peer composition in the classroom. More importantly, these results have direct implications for school tracking. Specifically, they suggest that an evaluation of tracking should focus more on the effects it has on teaching and curriculum, and less on whether some students are left disproportionately exposed to higher- or lower-ability students.

2 School Tracking in Thailand

Thailand has a 6-3-3 school system where students attend elementary school for 6 years, middle schools for 3 years, and then high school for 3 years. Typically, public schools only provide either primary education or secondary education. This means that the majority of Thai students have to start at a new school when they transition from primary education to secondary education in seventh grade.

At public middle schools, the practice of sorting students into classrooms based on ability, or 'tracking', is a common practice. The newly-enrolled seventh-grade students typically have to sit for a preliminary exam before the school year starts. Schools then use the results of this preliminary exam as a proxy for student ability and then sort students into classrooms based on the preliminary exam score. For example, in a school where there are 120 seventh-grade students, the 40 students

[2]For example, see Hoxby (2000), Lefgren (2004), Lavy and Schlosser (2011), Ohinata and Van Ours (2013), Sacerdote (2001), Zimmerman (2003), Carrell, Fullerton, and West (2009).

who scored the highest in the preliminary exam (rank 1-40) are normally sorted into class 1, the next 40 students (rank 41-80) are sorted into class 2, and the bottom 40 students (rank 81-120) are grouped together into class 3. Students who are assigned to the same class stay together in that class for at least a whole school year. Specifically, this means that they sit in the same classroom, follow the same timetable, and take all the same courses from the same teachers.

Importantly, one unique feature of the Thai schools in this paper is the fact that students in classrooms above and below the cutoff also take the same courses, follow the same curriculum, and take the same exams. And while the sets of teachers assigned to classrooms above and below the cutoffs might not always be completely identical, they are very similar. This is because teachers teach more than one classroom within a grade, which means that most classrooms above and below the cutoff have the same teacher for that subject.[3] As a result, the only thing changing at the classroom cutoff here is essentially the level of student ability in the classroom. This setting thus allows me to apply a regression discontinuity approach to identify the impact of solely peer quality in the tracking system. This is because comparable students who are just above and below the threshold are assigned to different classrooms that are the same in all aspects except for peer quality.

3 Data

The analyses in this paper use administrative data of students who were enrolled in the seventh grade in four public middle schools in Bangkok between 2013-2014 and 2016-2017. The data set consists of students' preliminary exam scores, class assignment, timetable, teachers assigned, GPA, and student characteristics, which include gender, height, weight, class size, birth order, and parents' marital status.

These four schools are all public secondary schools from the suburban area of Bangkok and were not chosen with any ex-ante presumptions. Rather, they were the only schools that kept

[3]I also account for any difference in teacher assignment by including teacher fixed effects.

complete records of the class sorting criteria and also allowed me access to the administrative data. For all schools, I checked each classroom's timetable to see if all the classrooms in the same grade, especially the classrooms just above and below the same cutoff, follow the same curriculum. I found that while it is true that students in all classrooms take the same core courses, the curriculum is more flexible for non-core subjects, such as physical education (PE). For example, there are instances where students in all classrooms take PE, but different classrooms take different PE courses. In other words, although all classes take PE, some classrooms have basketball, while some classrooms have volleyball. This difference is likely due to the fact that schools do not have enough equipment and teachers to allow all students to take the same non-core courses. Therefore, while students in all classrooms still follow the same curriculum and take the same required number of non-core courses in each semester, the non-core courses they take are sometimes different. I, therefore, limit my sample to only the cutoffs where the classrooms above and below follow the same identical courses, so that the only thing changing at the cutoff is student quality. As a result, my analysis sample consists of 10 cutoffs and 1,602 students.[4]

The main outcome of interest in this paper is seventh grade cumulative GPA. Importantly, GPA in Thai middle schools is based primarily on performance on multiple-choice exams for which there are no curves. Students who take the same course in the same school also take the same exact exams even when they are in different classrooms and are taught by different teachers. As a result, in contrast to other contexts, there is little scope for teacher subjectivity to affect student grades and GPA. Anecdotally, 10 percent of the final grades could be subject to teacher discretion, which is usually based on student attendance and attentiveness in class. Many students receive full points and very few students receive less than 5 percent out of this 10 percent portion. Since different schools in different school years could have different standards for GPA, I use standardized cumulative GPA instead of raw cumulative GPA. I standardize cumulative GPA by

[4]For one of the ten cutoffs, there are five classrooms below the cutoff and students who score below the cutoff are randomly assigned into one of the five classrooms. Out of these five classrooms, four classrooms follow the exact same curriculum as the classrooms above the cutoff, while one classroom does not have the same non-core courses as the classroom above the cutoff. There are 51 students in this particular classroom and I drop them from my sample. Since students are randomly assigned to this classroom, this should not affect results.

rescaling within each school and school year so that the mean of the standardized cumulative GPA is zero and the standard deviation is one.

Table 1 summarizes the characteristics of students in my data set. First, Column 1 reports the descriptive statistics for all the students in the data set used for the analysis. Columns 2-4 report the same statistics, but limit the sample to only students closer to the cutoff. Specifically, Column 2 reports the descriptive statistics of students whose preliminary exam score is within 20 points from the cutoff. Column 3 and Column 4 report the same statistics of students whose preliminary exam score is within 10 points and 5 points from the cutoff, respectively. Based on Table 1, the average preliminary exam score of the student in the sample is 48 percent. This is not surprising as the schools in the dataset are not selective schools. The average seventh grade cumulative GPA is 2.9. Columns 2-4 also show that 96 percent of students in the sample have a preliminary exam score within 20 points of the cutoff, while 65 percent are within 10 points and 40 percent are within 5 points.

4 Empirical Strategy

To disentangle the effects of peers from confounding factors, I apply a regression discontinuity design (RDD) that compares students just above and below the cutoff. The key assumption is that all other determinants of the outcomes except peer quality vary smoothly across the threshold.

Since I have multiple cutoffs, each with different cutoff scores, I follow the method used in Pop-Eleches and Urquiola (2013) and employ the stacked RDD method. Specifically, I first normalize the cutoffs using equation (1). Then I pool all normalized cutoffs together for the regression discontinuity analysis.

$$r_{ic} = prelim_i - cutoff\ score_c \tag{1}$$

In equation (1), the normalized preliminary exam score of student i from cutoff c is denoted by r_{ic}. $Prelim_i$ is student i's raw preliminary exam score and $cutoffscore_c$ denotes the cutoff score of cutoff c. Using equation (1), all the cutoff scores are recentered to zero. The normalized preliminary exam score (r_{ic}) indicates how far each student is from their associate cutoff as well as whether they are above or below the cutoff. The number is positive for those above the cutoff and negative for those below the cutoff. The formal regression discontinuity analysis in this paper then use the following standard regression discontinuity model:

$$Y_{ic} = \alpha_1 + \beta I[r_{ic} \geq 0] + \gamma_1 r_{ic} + \gamma_2 r_{ic} I[r_{ic} \geq 0] + \delta_c + u_{ic} \tag{2}$$

Where Y_{ic} is the outcome variable of student i at cutoff c. r_{ic} is the model's running variable, which is student i's normalized preliminary exam score. $I[r_{ic} \geq 0]$ is a binary variable indicating whether student i is above the cutoff. δ_c represents a full set of cutoff dummies. Importantly, the coefficient of interest here is β which indicates whether there is a discontinuity in the outcome (Y_{ic}) at the cutoff.

One important thing that should be noted here is that since each school could have multiple cutoffs in a school year, it is possible that some students are associated with two cutoffs at the same time. For example, from the example earlier where there are three classrooms in the seventh grade, students who are in class 2 are associated with two cutoffs: the one separating class 1 and class 2, and the one separating class 2 and class 3. The students who are associated with two cutoffs thus appear in the data set twice and have two different normalized preliminary scores calculated based on each of their two different cutoffs. Due to these repeated observations, I cluster standard errors at the individual level.

4.1 Test of Identification

As with any RD design, the key identification assumption here is that students just above and below the cutoff are comparable in the absence of treatment. I will be able to accurately estimate the impact of peer quality only if students just above and below the cutoff are comparable and the only things changing at the cutoff are their class assignment and the resulting peer quality. Under this assumption, any discontinuity in student achievement at the cutoff can be properly attributed to the increase in peer quality. In this section, I provide support for this approach by providing empirical evidence consistent with the identifying assumption.

To this end, I start by checking that students could not manipulate the cutoff. This is important because if some students could strategically place themselves just above the cutoff then it would mean that students just above and below the cutoff are fundamentally different. For example, one might worry if particularly motivated students were able to obtain scores just above the cutoff. Institutionally, there is no reason to believe that students would be able to manipulate their position relative to the cutoff. First and foremost, since the cutoff score was not known to the students before the preliminary exam, it would be difficult, if not impossible, for students to precisely predict where the cutoff will be and put in just the right amount of effort as to place themselves just above the cutoff. Moreover, there is also no retake of the preliminary exam. To provide further support for this institutional claim, I examine the distribution of students' normalized preliminary exam scores. If students could precisely manipulate their position relative to the cutoff, we would see a jump in the density of students at the cutoff. Figure 1 shows that the distribution of students' normalized preliminary score is smooth across the cutoff. The data are therefore consistent with my understanding of how students are assigned to classrooms and suggest no evidence of manipulation around the cutoff.

In addition, I also test whether the observable characteristics of students are smooth across the threshold. If the identifying assumption holds, all characteristics should be smooth across the cutoffs. If students could manipulate the threshold, we might observe a discontinuity at the cutoff for some characteristics. Here, I look at all the observable characteristics available in the data set including gender, weight, height, birth order, parents' marital status, and class size. In Figure 2, I show graphically that all characteristics are smooth across the cutoff. I then formally estimate the discontinuity of each covariate at the cutoff using the model described in the last section. The regression discontinuity estimates are reported in Table 2 and confirm that there is no statistically significant discontinuity in characteristics at the cutoff.

Additionally, rather than focusing on each of the characteristics individually, I also use these observable characteristics to predict seventh-grade cumulative GPA for each student. I then look at whether these predicted GPAs are smooth at the cutoff. The benefit of this method is that it allows me to attribute appropriate weight to each characteristic according to how much it contributes to student GPA. Figure 3 shows visually that there is no discontinuity in the predicted GPA at the cutoff. This again suggests that students just above and below the cutoff are comparable and that there is no manipulation of the threshold.

One limitation of the data is that I do not observe seventh grade cumulative GPA for roughly 14 percent of the students in my data set. This is because the schools did not provide me with the records of students who had transferred to another school or dropped out. This could potentially bias my estimates if there is selective attrition across the cutoff. To assess this, I test for a discontinuity in the probability of being observed with seventh grade cumulative GPA across the cutoff and show that there is no such discontinuity. Results are shown in Figure 1A and Table 1A in the Appendix. In addition, I also check for discontinuities in student characteristics and predicted GPAs again using only the students for whom I observe the main outcome, i.e. seventh grade cumulative GPA. The results hold and confirm that there is no difference in student characteristics at the cutoff. These results are shown in Figure 2A, Figure 3A, and Table 2A in the Appendix.

Based on all the evidence shown in this Section, I conclude that students observed in the sample on either side of the cutoff are comparable. This is consistent with the identifying assumption, and with the institutional background that suggests manipulation would be difficult, if not impossible, in this context. As a result, there is little reason to expect that student outcomes would be different on either side of the cutoff, absent the effect of being tracked with higher-ability peers.

5 Results

5.1 The Discontinuity in Classroom Assignment

First, I examine the first-stage relationship between students' normalized preliminary exam scores and their class assignments. Specifically, I examine how crossing the classroom cutoff affects students' probability of being in the higher-ability classroom. Panel 1 of Figure 4 shows visually that the probability of students being in the higher-ability classroom jumps from approximately 0 to 80 percent when they cross the cutoff. The reason why the compliance rate is not jumping from precisely 0 to 100 percent at the cutoff is that there are students who received special treatment and students who opted out of the assigned classrooms.[5]

I formally estimate and report the discontinuities in the probability of being tracked into higher-ability classrooms at the cutoff in Panel 1 of Table 3. The odd-numbered columns show the estimates from the regression without any controls, while the even-numbered columns show the estimates from the regression with controls for student characteristics. Columns 1-2 show the estimates from the regression using the full sample, while Columns 3-8 report the estimates from when the sample is only limited to students closer to the cutoff. The estimates reported in this panel range from 0.74-0.86 and all are statistically significant at conventional levels. In addition, across all bandwidths, the estimates change little as controls are added, consistent with the identifying assumption.

5.2 The Discontinuity in Peer Quality

Next, I turn my attention to peer quality. In this section, I examine whether crossing the cutoff and therefore having a higher chance of being in the higher-ability classroom is associated with higher quality peers. I measure each student's peer quality by calculating the average of their

[5]In one of the schools, students could choose to opt-out of their assigned classroom and enroll in the 'gifted' classroom if they could pay the higher tuition of the 'gifted' classroom. I leave the 51 students who were enrolled in the 'gifted' classroom in this school in the sample in order to avoid selection bias due to their exclusion, as the decision to switch could depend on which side of the threshold they were on. In Table 3A in the Appendix, I show that the decision to control or not control for these gifted classrooms in the regression does not affect my results.

classmates' standardized preliminary exam scores.[6]

Panel 2 of Figure 4 shows graphically that peer quality jumps by approximately 0.7 standard deviations at the classroom cutoff. The formal estimates are reported in Panel 2 of Table 3. They show that corresponding to the increase in the probability of being in the higher-ability classroom, peer quality jumps by 0.70-0.82 standard deviations at the cutoff. Again, my estimates are stable across bandwidth sizes and robust to the inclusion of student characteristic controls.

5.3 Reduced-Form Estimation: Effects on Seventh-Grade GPA

In the previous section, results indicate that crossing the cutoff is associated with an increase of approximately 0.70-0.82 standard deviations in peer quality. In this section, I examine whether this could, in turn, lead to an increase in academic performance, as measured by GPA. If it does, because peer quality is the only thing changing at the cutoff, it would suggest that crossing the cutoff increases student academic achievement through improvement in peer quality.

Figure 5, which plots the relationship between students' normalized preliminary score and standardized GPA, graphically shows this reduced-form relationship. From Figure 5, it is clear that there is no discontinuity in student GPA at the cutoff. This suggests that crossing the cutoff, and therefore having higher-quality peers, does not lead to better student outcomes.

I formally estimate the discontinuity in student GPA at the cutoff by estimating the model in equation (1) with standardized seventh-grade cumulative GPA as the outcome variable. The estimates are shown in Table 4. They are all statistically insignificant at conventional levels and range from -0.09 to -0.11 (Columns 1, 4, 7, 10). When I also include characteristic controls in my specification, across bandwidth sizes, the estimates change little. They are still statistically

[6] The standardized preliminary score of student j who is a 7th-grade student in school s in school year y is calculated using

$$standardized\ prelim_j = \frac{prelim_j - mean\ prelim_{sy}}{s.d.\ prelim_{sy}}$$

and i's peer quality is calculated using

$$peer\ quality_{ic} = peer\ quality_i = \frac{1}{n_{class(i)} - 1} \sum_{j \neq i, j \in class(i)} standardized\ prelim_j$$

insignificant at the conventional levels and range from -0.08 to -0.12 (Columns 2, 5, 8, 11). This suggests that the results are robust to the inclusion of controls and bandwidth sizes.

However, one might be concerned about teacher quality across classrooms. For instance, if the classrooms above the cutoff always get worse teachers, then my estimates of peer effects could be biassed downward. As mentioned before, institutionally, this should not be an issue as most of the teachers in the data set teach both the classrooms above and below the cutoff. Nevertheless, I address this issue empirically by adding teacher fixed effects to my specification. The estimates from this specification with teacher fixed effects become a little more negative and range from -0.11 to -0.19 (Columns 3, 6, 9, 12), but are still statistically insignificant at conventional levels. This suggests that if anything, students with higher ability peers may have access to higher-quality teachers, causing my unconditional estimates to be an upper bound. I emphasize, however, that the estimates without and with teacher fixed effects are not statistically different from each other.

In any case, since the estimates across specifications and bandwidth sizes are negative and statistically insignificant, the important thing we could take from the results is that being tracked into classrooms with higher-ability peers does not lead to significantly higher achievement for students. Importantly, the majority of the upper bounds of the 95 percent confidence intervals, which are also shown graphically in Figure 6, indicate that the effect of crossing the cutoff and therefore having higher-ability peers is not greater than 0.07 standard deviations.

5.4 2SLS Estimates

Next, in Table 5, I report local average treatment effect (LATE) estimates of being tracked into higher-ability classrooms using 2SLS. Intuitively, these estimates are the reduced-form estimates divided by the increase in the likelihood of attending the higher-ability classroom as shown in Panel 1 of Table 3. Estimates from Panel 1 of Table 5 indicate that peer quality increases by approximately 0.94 standard deviations when students are tracked into higher-ability classrooms. At the same time, Panel 2 of Table 5 reports that being tracked into higher-ability classrooms and therefore having peers that are 0.94 standard deviations better is associated with a statistically insignificant 0.10-0.16 standard deviation decrease in student GPA. Additionally, Panel 3 of Table 5 rescales the estimates and shows that an increase of one standard deviation in classroom peer quality results in a statistically insignificant decrease in student GPA of 0.10-0.18 standard deviations.

In addition, Figure 7 plots the LATE estimates of being in higher-ability classrooms on student achievement (GPA) along with their 95 percent confidence intervals across bandwidth sizes. We can see that the estimates are all negative, statistically insignificant, and relatively stable across bandwidth sizes. Importantly, more than 80 percent of the upper bound estimates across bandwidth sizes are smaller than 0.08 standard deviations. This enables me to rule out any positive effects bigger than 0.08 standard deviations.

While Figure 7 plots the LATE estimates of being in the high-ability classrooms which are associated with an increase of 0.94 standard deviations in peer quality, Figure 8 shows the LATE estimates of an increase of one standard deviation in peer quality across bandwidth sizes. Because of the large first-stage discontinuity, the estimates in Figure 8 are very similar to those in Figure 7. They are all negative and statistically insignificant and the upper bounds suggest that an increase of one standard deviation in classroom peer quality could not lead to an increase in student achievement that is larger than 0.08 standard deviations. To summarize, results from 2SLS estimations indicate that being tracked into better classrooms is associated with an increase of 0.94 standard deviations in peer quality. However, that increase in peer quality does not lead to positive effects on GPA, as point estimates are negative and I am able to rule out positive effects larger than 0.08 standard deviations.

6 Discussion

The absence of a positive effect for students tracked into classrooms with significantly higher-achieving peers seems puzzling. There are multiple possible interpretations of this finding. One is that perhaps there are positive effects for one group that are offset by negative or null effects for another group. To investigate this possibility at least as it relates to perhaps the most salient difference—student gender—I look at peer effects on male and female students separately. The results, shown visually in Figure 4A in the Appendix, suggest that the impacts of peer quality are similar for both genders and that there are no positive peer effects for either male or female students. And while some of the formal estimates, shown in Table 4A in the Appendix, are statistically significant, they are marginally significant and not robust. Importantly, the underlying plots shown in Figure 4A show little evidence that there is any discontinuity in student GPA at the cutoff for either gender. Therefore, it seems highly unlikely that I do not detect effects because of this reason.

While the most obvious interpretation of these findings is that exposure to higher-achieving peers does not benefit students, it is also possible that any benefits from that exposure are offset by other differences. For example, when a student is tracked into a classroom with higher-ability peers, he/she also automatically becomes a small fish in a large pond and has a lower rank in the classroom. This means that the impact of being tracked into a higher-ability classroom captures the net effect of increased exposure to high-ability peers, but also lower relative rank. Murphy and Weinhardt (2018) have looked into the effects of ordinal rank on student achievement and found large effects. These effects could potentially offset any positive effects from exposure to higher-achieving peers. I note that my setting is not unique in this sense; any policy that increases one's exposure to higher-achieving peers will also mechanically lower rank. In theory, I could untangle the two effects by looking at the heterogeneous effects across cutoffs. For example, I could compare the estimates at cutoffs with big increases in peer quality to the estimates from cutoffs with a small increase in peer quality, but both of which include similar effects on rank. Unfortunately, I do not have enough data and heterogeneity across cutoffs to do so in a constructive way. As a result, in this paper, I do not attempt to separate the two effects, but instead identify the reduced-form policy-relevant effect of being tracked into a classroom with higher- ability peers.

Additionally, the change in peer quality could also affect students through the change in teacher behavior. Specifically, teachers might tailor their instruction to the quality of the students in each classroom and therefore teach students in the high- and low- ability classrooms differently. If that is the case, it is possible that the change in teacher instruction affects students in the way that it offsets any positive effects from exposure to high-ability peers. In this paper, since I do not observe or have information on teacher instruction, I do not attempt to separate these indirect peer effects from the direct effects of having high-achieving peers. Instead, as stated earlier I focus on identifying the reduced-form policy-relevant effect of being tracked into a classroom with higher-ability peers.

7 Conclusion

This paper estimates the impacts of being tracked into classrooms with higher-achieving peers on student achievement using administrative data from public middle schools in Thailand. Using an RDD approach, reduced-form results show that crossing the classroom cutoff is associated with a large increase of approximately 80 percentage points in the likelihood of being assigned to the higher-ability classroom. This, in turn, translates to an increase of 0.7-0.8 standard deviations in peer quality at the cutoff. However, the increase in peer quality at the cutoff does not lead to an increase in student achievement as the seventh-grade cumulative GPAs remain smooth across the cutoff. Two-stage least squares estimates indicate that barely being in a higher-ability classroom is associated with a 0.94 standard deviation increase in peer quality, and results in a statistically insignificant 0.10-0.16 standard deviation reduction in student GPA.

Importantly, my results are robust to bandwidth size and the inclusion of student characteristic controls and teacher fixed effects. In addition, upper bound estimates also allow me to rule out positive peer effects of modest sizes. Specifically, my upper-bound estimates indicate that the effects of a significant increase of one standard deviation in peer quality on student GPA, at least in Asian contexts similar to this, could not be larger than 0.08 standard deviations.

These results are in line with the findings in Duflo, Dupas, and Kremer (2011), which found that an increase of one standard deviation in peer quality in classrooms in tracking schools leads to no statistically significant increase in student achievement. However, the strength of this paper is that I am able to rule out effects larger than 0.08 standard deviations, while they were only able to rule out effects larger than 0.21 standard deviations. My findings also enable me to speak to the literature on the effectiveness of tracked classrooms, such as gifted and talented programs, in general. My result complement the findings in Bui, Craig, and Imberman (2014); Card and Giuliano (2016); Cohodes (2020) which estimated few and insignificant test score effects of special classrooms. And while Booij, Haan, and Plug (2017) found effects of 0.2 standard deviations of a gifted and talented program in the Netherlands, my results suggest that exposure to significantly higher-quality peers could not increase student achievement more than 0.08 standard deviations. Thus, it seems likely that a large portion of the positive effects found in Booij, Haan, and Plug (2017) are the results of specific features of the program.

In addition, my estimates also rule out effects that are small relative to previous papers on the benefits of attending higher-quality schools, which also have better peers. For example, Pop-Eleches and Urquiola (2013) studied Romanian secondary schools and found that attending higher-quality schools where peers are on average 0.1 standard deviations higher in quality results in an increase of 0.02-0.10 standard deviation increase in high school exit exam. Given the results in this paper indicate that a one standard deviation increase in peer quality could not lead to an increase of more than 0.08 standard deviations in student achievement, results here suggest that less than half of the effects found in Pop-Eleches and Urquiola (2013) could be attributed to the increase in peer quality in better schools. In particular, my results complement the findings of Hoekstra, Mouganie, and Wang (2018) and Jackson (2013) that the returns to high school quality are likely the result of other features of the schools other than peer quality and that peer quality explains very little of those returns.

More generally, the results of this study suggest that parents and policymakers should perhaps focus less on peer quality when making decisions as to how to best improve educational outcomes for children. Additionally, an equivalent way of interpreting the results is that being tracked into lower-ability classrooms and therefore being exposed to lower-ability peers does not result in lower student achievement. The results suggest that at least in this context, concerns that tracking systems might disproportionately harm students tracked into lower-ability classrooms seem overemphasized. Rather, future work on tracking should focus more on the effects it has on teaching and curriculum, and less on whether some students are left disproportionately exposed to higher- or lower-ability students.

References

Abdulkadiroğlu, A., J. Angrist, and P. Pathak (2014). The elite illusion: Achievement effects at boston and new york exam schools. *Econometrica 82*(1), 137–196.

Abdulkadiroglu, A., P. A. Pathak, J. Schellenberg, and C. R. Walters (2017). Do parents value school effectiveness? Technical report, National Bureau of Economic Research.

Allensworth, E. M., P. T. Moore, L. Sartain, and M. de la Torre (2017). The educational benefits of attending higher performing schools: Evidence from chicago high schools. *Educational Evaluation and Policy Analysis 39*(2), 175–197.

Booij, A. S., F. Haan, and E. Plug (2017). Can gifted and talented education raise the academic achievement of all high-achieving students?

Bui, S. A., S. G. Craig, and S. A. Imberman (2014). Is gifted education a bright idea? assessing the impact of gifted and talented programs on students. *American Economic Journal: Economic Policy 6*(3), 30–62.

Card, D. and L. Giuliano (2016). Can tracking raise the test scores of high-ability minority students? *American Economic Review 106*(10), 2783–2816.

Carrell, S. E., R. L. Fullerton, and J. E. West (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics 27*(3), 439–464.

Clark, D. (2010). Selective schools and academic achievement. *The BE Journal of Economic Analysis & Policy 10*(1).

Cohodes, S. R. (2020). The long-run impacts of specialized programming for high-achieving students. *American Economic Journal: Economic Policy 12*(1), 127–66.

Dee, T. and X. Lan (2015). The achievement and course-taking effects of magnet schools: Regression-discontinuity evidence from urban china. *Economics of Education Review 47*, 128–142.

Dobbie, W. and R. G. Fryer Jr (2014). The impact of attending a school with high-achieving peers: Evidence from the new york city exam schools. *American Economic Journal: Applied Economics 6*(3), 58–75.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review 101*(5), 1739–74.

Hoekstra, M., P. Mouganie, and Y. Wang (2018). Peer quality and the academic benefits to attending better schools. *Journal of Labor Economics 36*(4), 841–884.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research.

Jackson, C. K. (2013). Can higher-achieving peers explain the benefits to attending selective schools? evidence from trinidad and tobago. *Journal of Public Economics 108*, 63–77.

Lavy, V. and A. Schlosser (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics 3*(2), 1–33.

Lefgren, L. (2004). Educational peer effects and the chicago public schools. *Journal of urban Economics 56*(2), 169–191.

Loveless, T. (2013). The resurgence of ability grouping and persistence of tracking. *Part II of the 2013 Brown Center Report on American Education, the Brookings Institution.[2] 1*.

Lucas, A. M. and I. M. Mbiti (2014). Effects of school quality on student achievement: Discontinuity evidence from kenya. *American Economic Journal: Applied Economics 6*(3), 234–63.

Murphy, R. and F. Weinhardt (2018). Top of the class: The importance of ordinal rank. Technical report, National Bureau of Economic Research.

OECD (2013). Pisa 2012 results: What makes schools successful? resources, policies and practices (volume iv).

Ohinata, A. and J. C. Van Ours (2013). How immigrant children affect the academic achievement of native dutch children. *The Economic Journal 123*(570), F308–F331.

Park, A., X. S. Shi, C.-T. Hsieh, and X. An (2009). Does school quality matter?: Evidence from a natural experiment in rural china.

Pop-Eleches, C. and M. Urquiola (2013). Going to a better school: Effects and behavioral responses. *American Economic Review 103*(4), 1289–1324.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics 116*(2), 681–704.

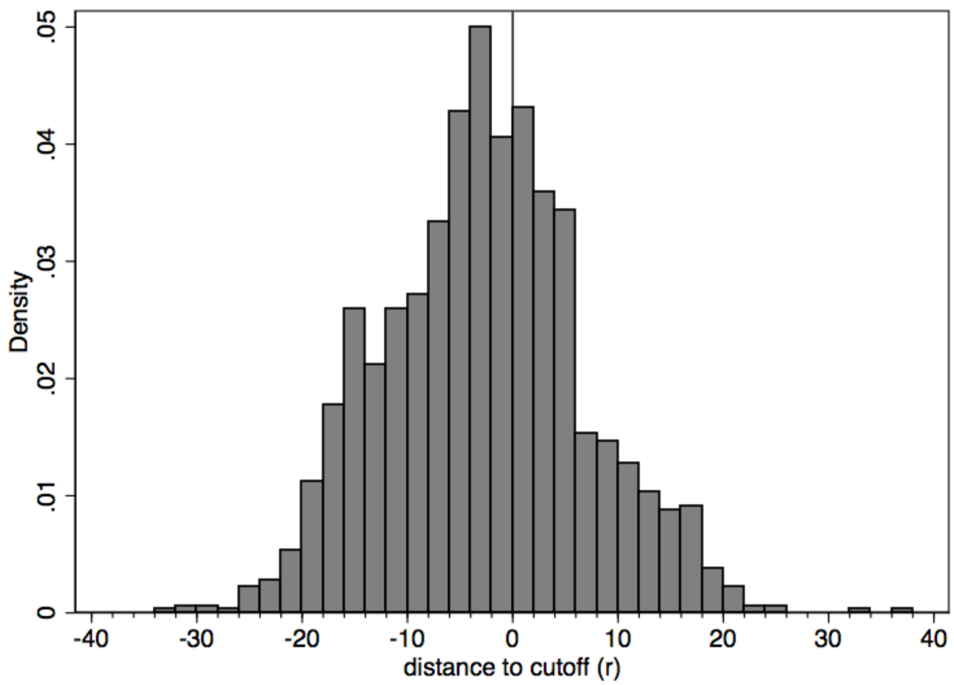
Shi, Y. (2019). Who benefits from selective education? evidence from elite boarding school admissions. *Economics of Education Review*, 101907.

Vardardottir, A. (2013). Peer effects and academic achievement: a regression discontinuity approach. *Economics of Education review 36*, 108–121.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics 85*(1), 9–23.

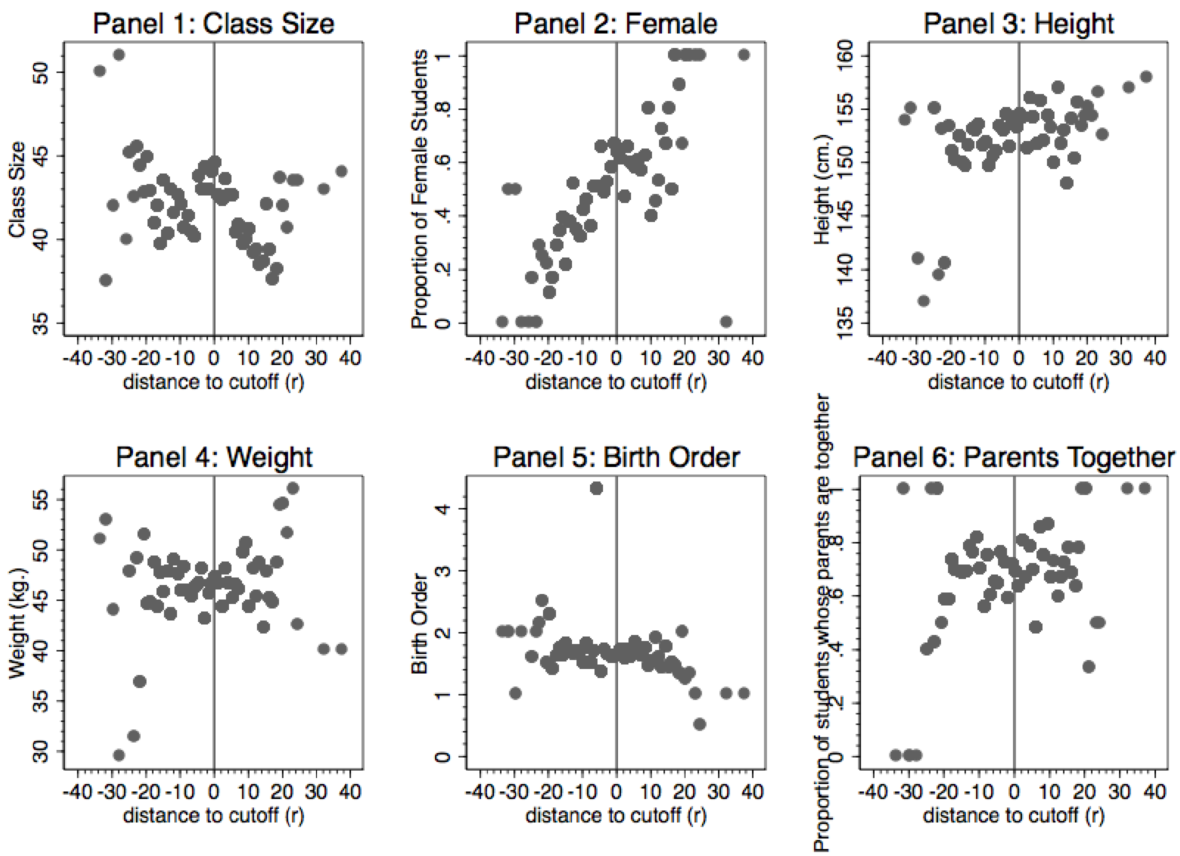
8 Figures & Tables

Figure 1: Histogram of running variable



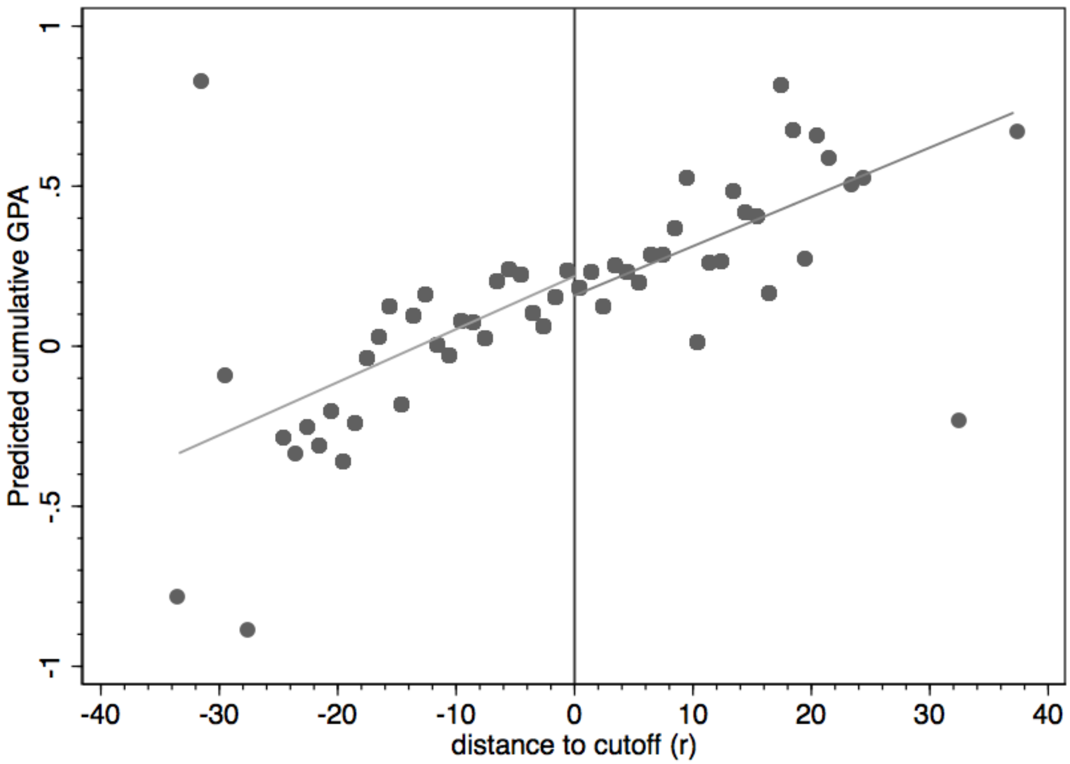
Notes: This figure shows the density of the running variable. The running variable used here is the distance to cutoff, i.e. how far each student's preliminary score is from the cutoff. The number is positive if they score above the cutoff, and negative if they score below the cutoff.

Figure 2: Student characteristics across cutoff



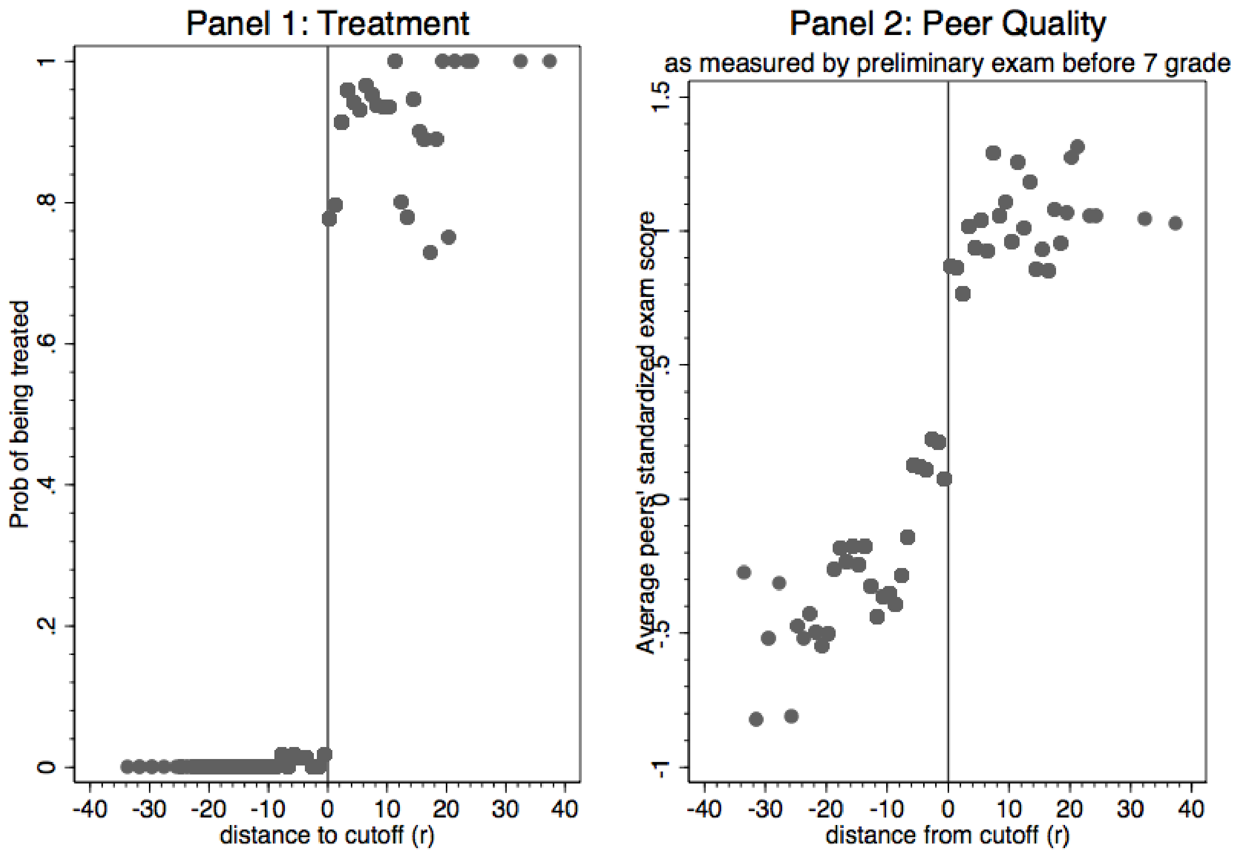
Notes: This figure plots student characteristics across the cutoff. This figure corresponds to Table 2.

Figure 3: Predicted 7th grade cumulative GPA based on student characteristics



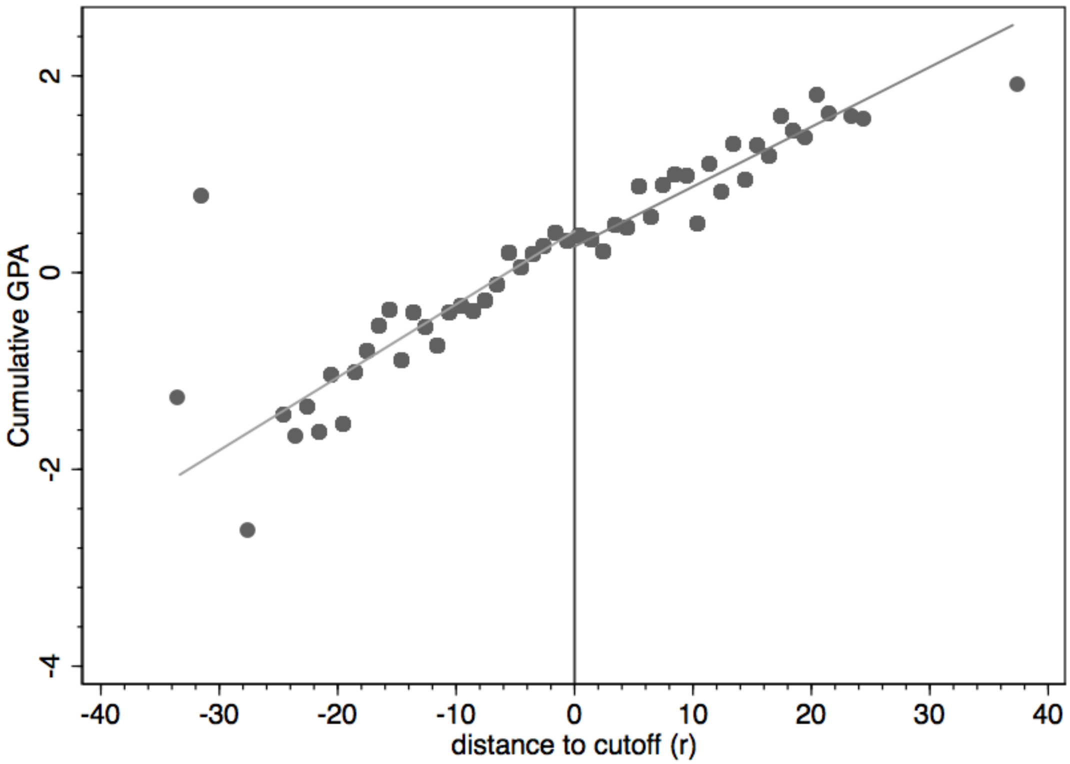
Notes: This figure plots predicted cumulative GPA across the cutoff. Predicted cumulative GPAs are based on the regression of cumulative GPA on student characteristics.

Figure 4: Probability of being in the better classroom and peer quality across cutoff



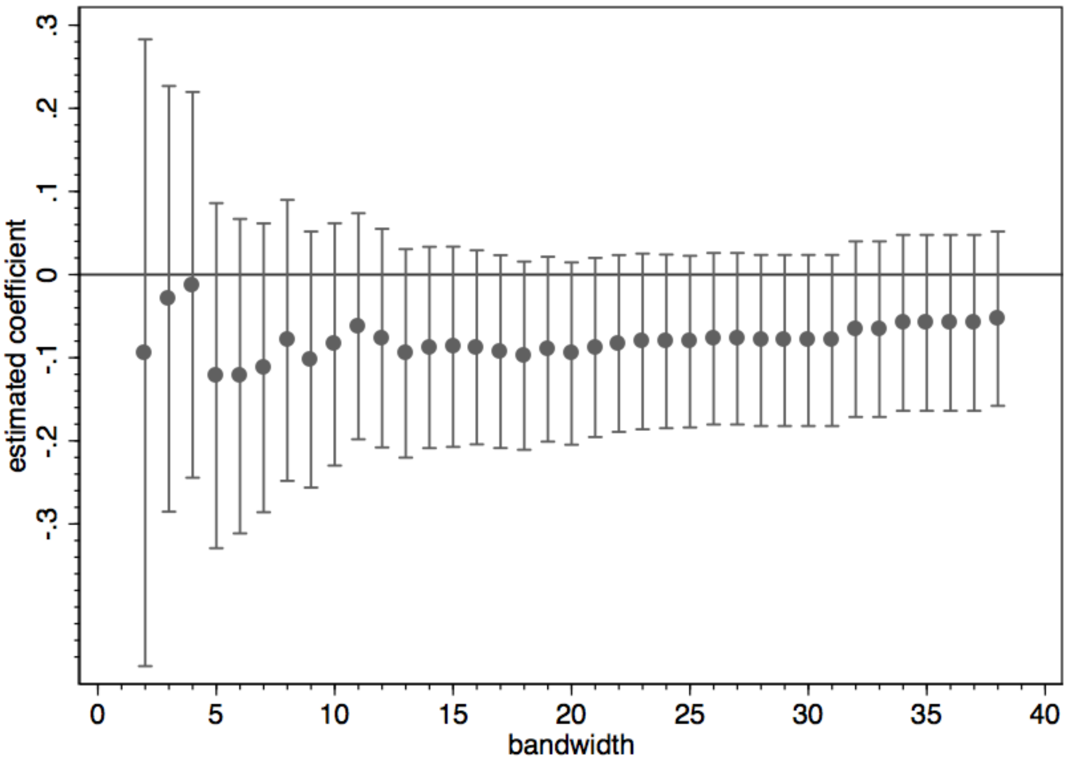
Notes: This figure plots the probability of being assigned to higher-ability classrooms and peer quality across the cutoff. Treatment indicates assignment to the higher-ability classroom. Peer quality is measured by the average of classroom peers' standardized preliminary exam score.

Figure 5: Standardized seventh grade cumulative GPA across cutoff



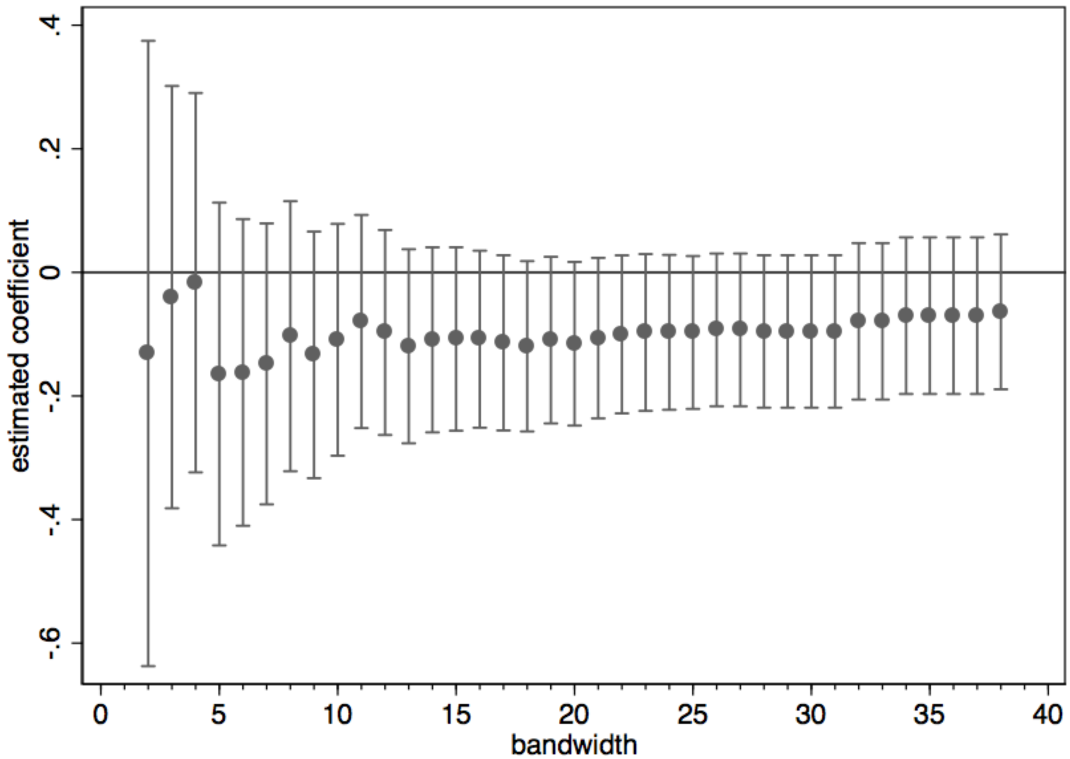
Notes: This figure shows the reduced-form relationship between students' normalized preliminary score (distance to cutoff) and standardized GPA.

Figure 6: Reduced-form estimates using different bandwidth sizes



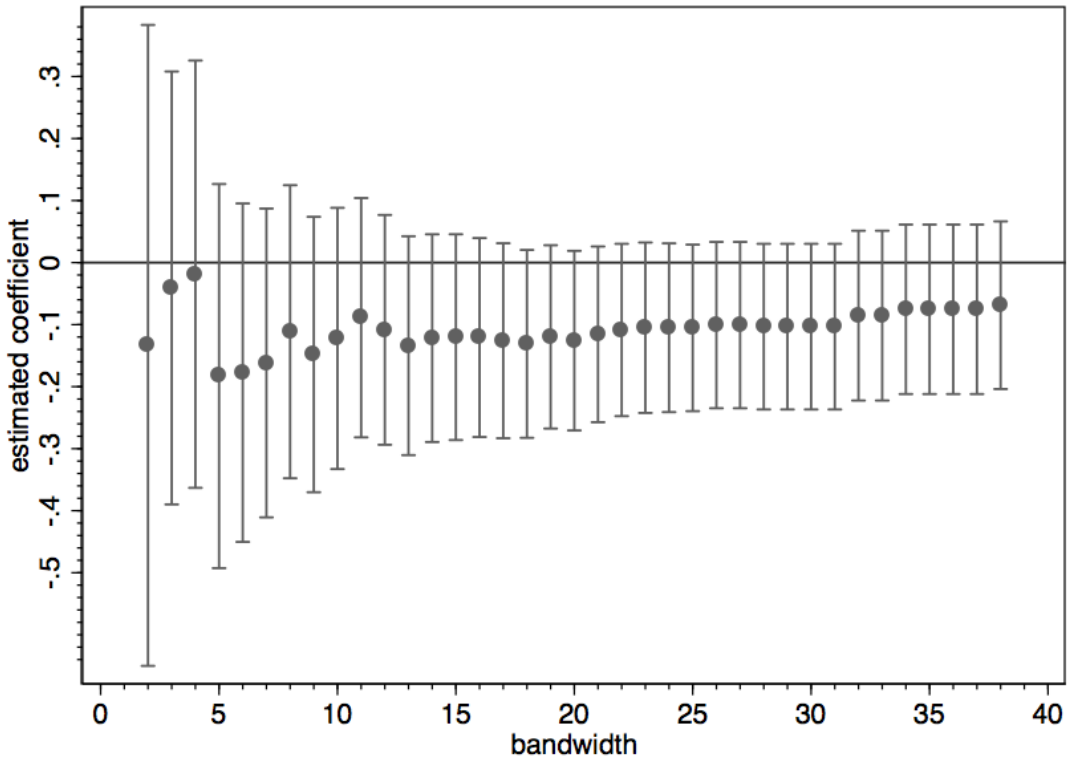
Notes: This figure shows the estimates from regressions using different bandwidth sizes. Estimates are from specification with controls for student characteristics

Figure 7: LATE estimates of being tracked into higher-ability classrooms on student GPA across bandwidth



Notes: This figure shows the estimates from regressions using different bandwidth sizes. Estimates are from specification with controls for student characteristics

Figure 8: LATE estimates of an increase of one s.d. in peer quality on student GPA across bandwidth



Notes: This figure shows the estimates from regressions using different bandwidth sizes. Estimates are from specification with controls for student characteristics

Table 1: Summary statistics

	(1) full sample	(2) $-20 < r < 20$	(3) $-10 < r < 10$	(4) $-5 < r < 5$
preliminary score	47.79	48.31	50.70	51.89
	(12.52)	(11.71)	(8.764)	(7.100)
distance to cutoff (r)	-2.971	-2.593	-0.952	-0.297
	(9.535)	(8.686)	(4.796)	(2.791)
class size	42.12	42.07	42.54	43.38
	(6.531)	(6.608)	(7.249)	(7.599)
female	0.511	0.517	0.563	0.585
weight (kg)	46.47	46.49	46.47	46.24
	(10.50)	(10.36)	(10.99)	(11.31)
height (cm)	152.6	152.7	153.1	153.6
	(9.045)	(8.974)	(8.775)	(8.788)
birth order	1.744	1.743	1.787	1.624
	(4.275)	(4.346)	(5.208)	(0.821)
Parents are together	0.697	0.701	0.697	0.711
7th-grade cumulative GPA	2.904	2.912	2.962	2.972
	(0.600)	(0.586)	(0.530)	(0.518)
standardized 7-th grade cumulative GPA	0.164	0.178	0.284	0.312
	(0.996)	(0.973)	(0.873)	(0.852)
Observations	1602	1543	1050	660

mean coefficients; sd in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Regression discontinuity estimates of student characteristics

	(1) full sample	(2) $-30 < r < 30$	(3) $-20 < r < 20$	(4) $-10 < r < 10$	(5) $-5 < r < 5$
Class size	0.3427*	0.3958**	0.4903**	0.4193	0.3842
	(0.1871)	(0.1909)	(0.1978)	(0.2637)	(0.3737)
Female	-0.03921	-0.05341	-0.04857	-0.02265	0.004405
	(0.03833)	(0.03885)	(0.04147)	(0.05466)	(0.07627)
Weight (kg)	-0.09691	-0.2505	0.06910	0.04397	1.2388
	(0.8806)	(0.9108)	(0.9152)	(1.3042)	(1.8857)
Height (cm)	0.1375	0.2473	0.3475	-0.4088	0.8303
	(0.7188)	(0.7425)	(0.7836)	(0.9714)	(1.3944)
Birth Order	-0.1084	-0.1020	-0.1390	0.1620	-0.03075
	(0.2663)	(0.2618)	(0.2494)	(0.1822)	(0.1352)
Parents together	0.008322	0.02086	0.02942	-0.0001297	-0.004345
	(0.03786)	(0.03885)	(0.04094)	(0.05367)	(0.07503)

Parentheses contain standard errors, clustered at individual level.

All regressions used rectangular kernel.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Regression Discontinuity Estimates for Treatment (First Stage)

	full sample		$-30 < r < 30$		$-20 < r < 20$		$-10 < r < 10$		$-5 < r < 5$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel 1. Probability of being in the higher-ability classroom										
Preliminary score above or at cutoff ($r \geq 0$)	0.855***	0.834***	0.855***	0.831***	0.852***	0.823***	0.796***	0.770***	0.763***	0.739***
	(0.0211)	(0.0190)	(0.0217)	(0.0195)	(0.0226)	(0.0203)	(0.0298)	(0.0275)	(0.0414)	(0.0387)
N	1602	1542	1595	1536	1543	1489	1050	1023	660	643
Panel 2. Peer Quality: Average standardized preliminary exam score of peers										
Preliminary score above or at cutoff ($r \geq 0$)	0.815***	0.794***	0.810***	0.789***	0.797***	0.774***	0.732***	0.712***	0.713***	0.698***
	(0.0256)	(0.0259)	(0.0260)	(0.0264)	(0.0267)	(0.0272)	(0.0342)	(0.0346)	(0.0488)	(0.0489)
N	1602	1542	1595	1536	1543	1489	1050	1023	660	643
Controls										
Cutoff fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Student characteristics		Y		Y		Y		Y		Y

Student characteristics include classsize, gender, height, weight, birth order, parents' relationship status.

Parentheses contain standard errors, clustered at individual level.

All regressions use rectangular kernel.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Reduced-form estimates

	$-30 < r < 30$			$-20 < r < 20$			$-10 < r < 10$			$-5 < r < 5$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Preliminary score above or at cutoff ($r \geq 0$)	-0.0908	-0.0793	-0.169	-0.102*	-0.0950*	-0.182	-0.0946	-0.0841	-0.189	-0.114	-0.122	-0.114
	(0.0577)	(0.0525)	(0.111)	(0.0613)	(0.0559)	(0.112)	(0.0831)	(0.0744)	(0.121)	(0.121)	(0.106)	(0.132)
Controls												
Cutoff fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Student characteristics		Y	Y		Y	Y		Y	Y		Y	Y
Teacher fixed effects			Y			Y			Y			Y
N	1366	1362	1362	1331	1328	1328	949	947	947	598	597	597

Student characteristics include classsize, gender, height, weight, birth order, parents' relationship status.

Parentheses contain standard errors, clustered at individual level.

All regressions use rectangular kernel.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: 2SLS estimates

	(1)	(2)	(3)	(4)
	$-30 < r < 30$	$-20 < r < 20$	$-10 < r < 10$	$-5 < r < 5$
Panel 1. Peer quality				
Being tracked into higher-ability classroom	0.9496***	0.9406***	0.9246***	0.9439***
	(0.02131)	(0.02199)	(0.02933)	(0.04030)
N	1536	1489	1023	643
Panel 2. Standardized 7th grade cumulative GPA				
Being tracked into higher-ability classroom	-0.09562	-0.1155*	-0.1091	-0.1645
	(0.06291)	(0.06756)	(0.09570)	(0.1415)
N	1362	1328	947	597
Panel 3. Standardized 7th grade cumulative GPA				
Peer quality increases by 1 s.d.	-0.1033	-0.1259*	-0.1223	-0.1830
	(0.06812)	(0.07386)	(0.1074)	(0.1580)
N	1362	1328	947	597
Controls				
Cutoff fixed effects	Y	Y	Y	Y
Student characteristics	Y	Y	Y	Y

Student characteristics include classsize, gender, height, weight, birth order, parents' relationship status.

Parentheses contain standard errors, clustered at individual level.

All regressions use rectangular kernel.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

9 Appendix

Figure 1A: Observability of seventh grade cumulative GPA across cutoff

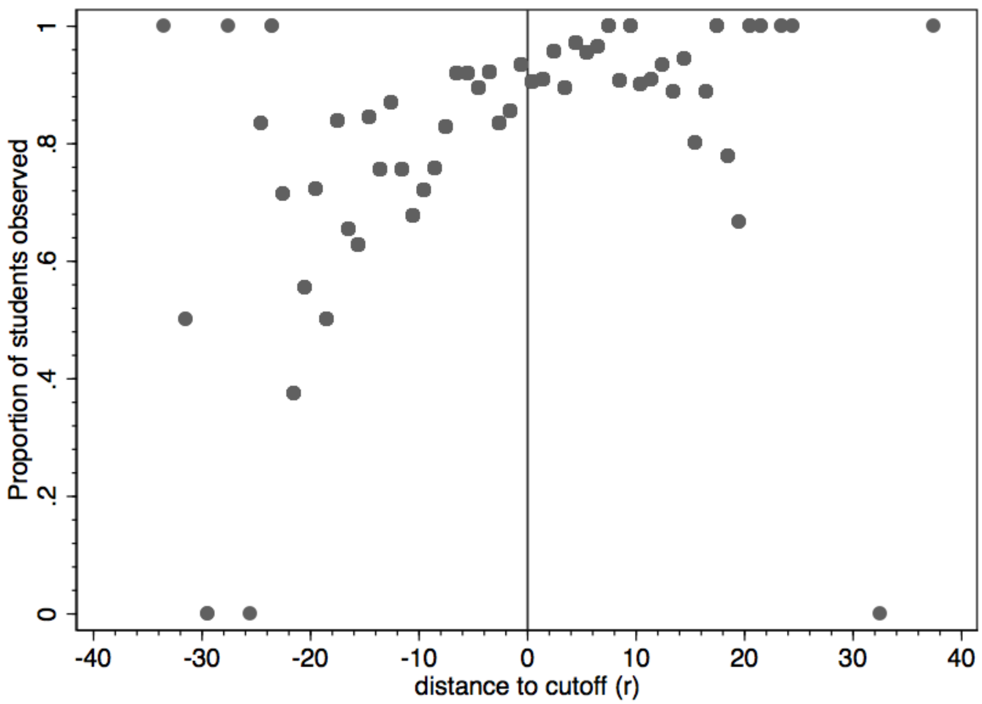


Figure 2A:Student characteristics across cutoff – only students whose seventh grade cumulative GPA is observed

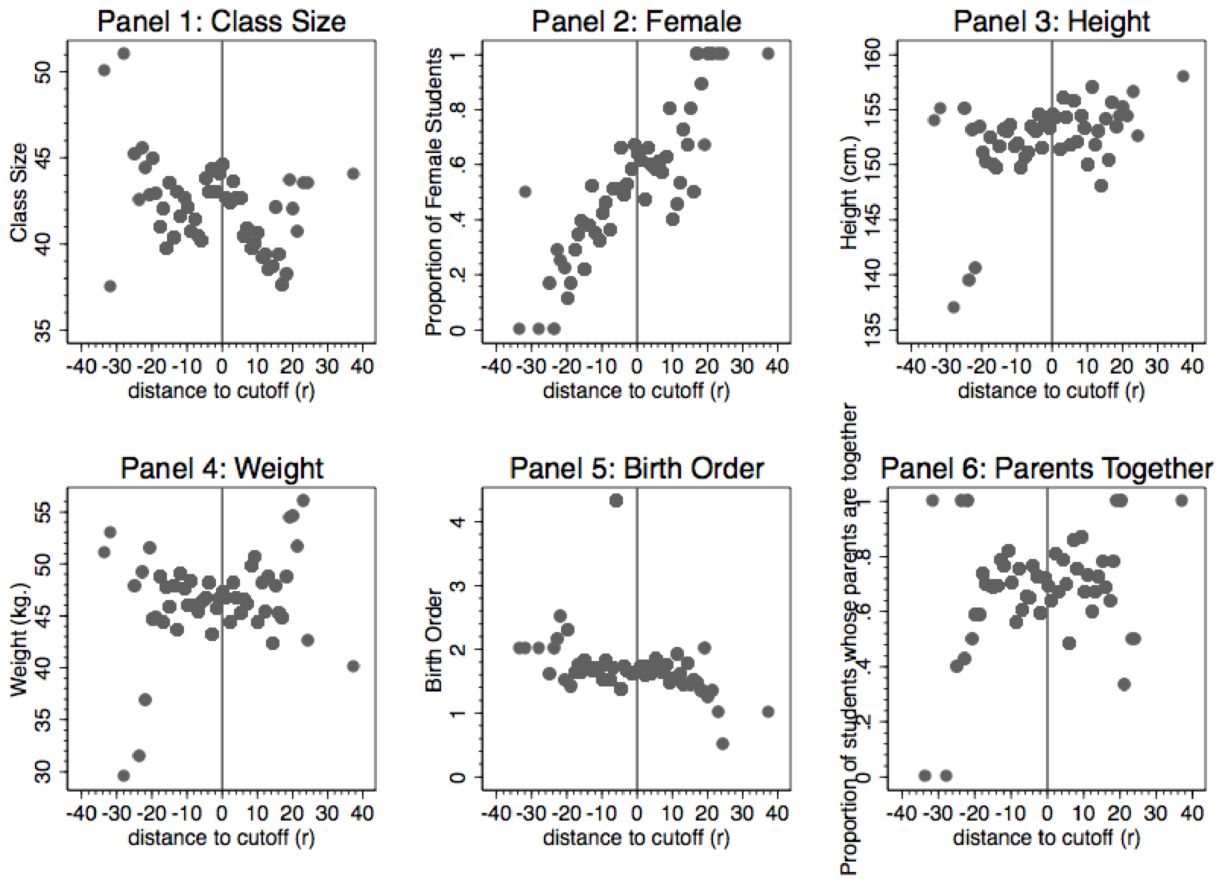


Figure 3A: Predicted 7th-grade cumulative GPA based on student characteristics – only students whose seventh grade cumulative GPA is observed

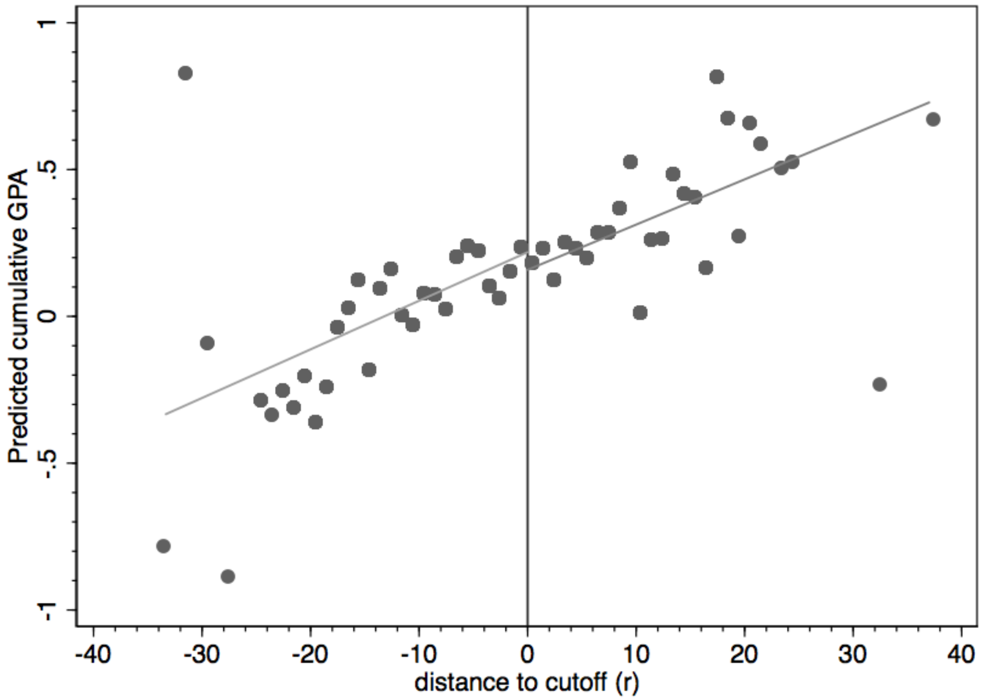


Figure 4A: 7th-grade cumulative GPA across cutoff by gender

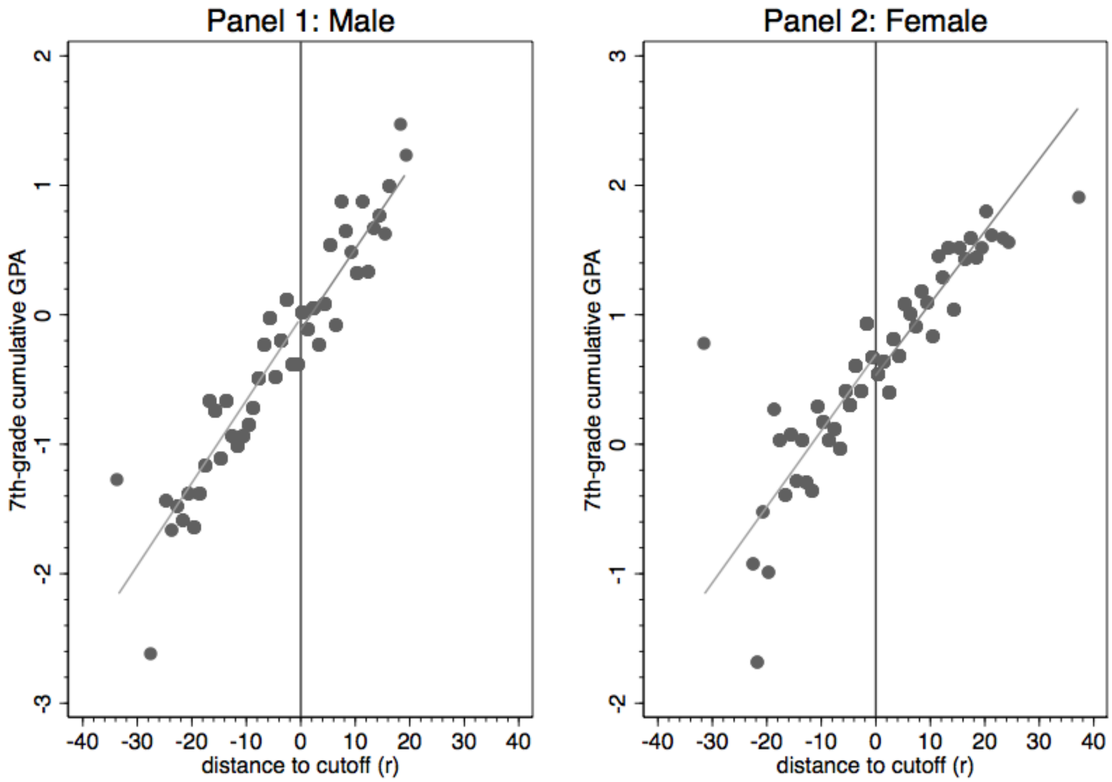


Table 1A: Regression discontinuity estimations of observability of seventh grade cumulative GPA

	$-30 < r < 30$		$-20 < r < 20$		$-10 < r < 10$		$-5 < r < 5$	
	observed	observed	observed	observed	observed	observed	observed	observed
above or at cutoff ($r \geq 0$)	0.009454 (0.02570)	0.0007272 (0.02268)	0.01702 (0.02765)	0.005156 (0.02410)	0.01856 (0.03324)	0.009819 (0.02987)	-0.001476 (0.04326)	0.009649 (0.03929)
Controls								
Cutoff fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Student characteristics	Y	Y	Y	Y	Y	Y	Y	Y
Teacher fixed effects		Y		Y		Y		Y
N	1595	1536	1543	1489	1050	1023	660	643

Student characteristics include classsize, gender, height, weight, birth order, parents' relationship status.

Parentheses contain standard errors, clustered at individual level.

All regressions use rectangular kernel.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2A: Regression discontinuity estimates of student characteristics– only students whose seventh grade GPA is observed

	(1) full sample	(2) $-30 < r < 30$	(3) $-20 < r < 20$	(4) $-10 < r < 10$	(5) $-5 < r < 5$
Class size	0.4230*	0.4534**	0.5730**	0.5428	0.4877
	(0.2057)	(0.2078)	(0.2158)	(0.2865)	(0.4065)
Female	-0.02082	-0.02440	-0.008761	0.004948	0.03803
	(0.04082)	(0.04138)	(0.04423)	(0.05771)	(0.08024)
Weight (kg)	-0.01964	-0.1493	0.07158	0.03222	1.3482
	(0.9425)	(0.9652)	(0.9685)	(1.3726)	(1.9759)
Height (cm)	0.4673	0.4777	0.5954	-0.2890	1.2295
	(0.7716)	(0.7897)	(0.8289)	(1.0105)	(1.4292)
Birth Order	-0.1468	-0.1479	-0.1559	0.1387	-0.03113
	(0.2982)	(0.2946)	(0.2800)	(0.1842)	(0.1398)
Parents together	-0.001599	0.003263	0.008496	-0.005200	-0.01253
	(0.03912)	(0.03984)	(0.04182)	(0.05457)	(0.07642)

Parentheses contain standard errors, clustered at individual level.

All regressions used rectangular kernel.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3A: Reduced-form estimates (including vs. not including gifted classrooms)

As students in one of my schools could opt out of the assigned classroom and enroll in the gifted classroom after taking the preliminary exam, dropping students in the gifted classrooms from this school might incur selection issue. I, therefore, chose to keep them in my sample. This table shows that the decision to control or not control for the gifted classroom does not affect my results.

	$-30 < r < 30$			$-20 < r < 20$			$-10 < r < 10$			$-5 < r < 5$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel 1: Without controls for gifted classrooms												
Preliminary score above or at cutoff ($r \geq 0$)	-0.0908	-0.0793	-0.169	-0.102*	-0.0950*	-0.182	-0.0946	-0.0841	-0.189	-0.114	-0.122	-0.114
	(0.0577)	(0.0525)	(0.111)	(0.0613)	(0.0559)	(0.112)	(0.0831)	(0.0744)	(0.121)	(0.121)	(0.106)	(0.132)
N	1366	1362	1362	1331	1328	1328	949	947	947	598	597	597
Panel 2: With control for gifted classrooms												
Preliminary score above or at cutoff ($r \geq 0$)	-0.0923	-0.0900	-0.169	-0.103*	-0.106*	-0.182	-0.0982	-0.0986	-0.189	-0.112	-0.125	-0.114
	(0.0576)	(0.0534)	(0.111)	(0.0612)	(0.0570)	(0.112)	(0.0831)	(0.0759)	(0.121)	(0.121)	(0.106)	(0.132)
N	1366	1362	1362	1331	1328	1328	949	947	947	598	597	597
Controls												
Cutoff fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Student characteristics		Y	Y		Y	Y		Y	Y		Y	Y
Teacher fixed effects			Y			Y			Y			Y

Student characteristics include classsize, gender, height, weight, birth order, parents' relationship status.

Parentheses contain standard errors, clustered at individual level.

All regressions use rectangular kernel.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A: Reduced-form estimates by gender

	$-30 < r < 30$		$-20 < r < 20$		$-10 < r < 10$		$-5 < r < 5$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel 1. Male								
Preliminary score	-0.0861	-0.314*	-0.111	-0.344*	-0.0286	-0.312	0.0850	-0.0287
above or at cutoff	(0.0908)	(0.178)	(0.0936)	(0.180)	(0.136)	(0.203)	(0.209)	(0.256)
N	635	635	616	616	407	407	245	245
Panel 2. Female								
Preliminary score	-0.0578	-0.164	-0.102	-0.189	-0.143*	-0.188	-0.263**	-0.259
above or at cutoff	(0.0707)	(0.127)	(0.0703)	(0.124)	(0.0847)	(0.137)	(0.115)	(0.164)
N	730	730	712	712	540	540	352	352
Controls								
Cutoff fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Student characteristics	Y	Y	Y	Y	Y	Y	Y	Y
Teacher fixed effects		Y		Y		Y		Y

Standard errors in parentheses

* p¡0.10, ** p¡0.05, *** p¡0.010