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

Meradee Tangvatcharapong *

Dec 29, 2018

Abstract

A common educational practice around the world is to track students into classrooms based on ability. However, despite the popularity of tracking, relatively few papers directly identify the impact of being tracked into classrooms with higher or lower peer ability. This paper estimates the impact of being tracked into a classroom with higher ability peers by using data from public middle schools in Thailand, where students are tracked into classrooms based on a preliminary exam taken before 7th grade. Importantly, all teachers, curriculum, and textbooks are identical throughout classrooms. To distinguish the impact of peers from confounding factors due to selection, I apply a regression discontinuity design (RDD) that compares the academic outcomes of students just above and below the threshold. Results indicate that significant increases in peer quality do not improve student GPA. This suggests that any gains due to tracking, at least in Asian contexts similar to this, are likely due to factors other than peer quality, such as curriculum or teacher quality.

*Department of Economics, Texas A&M University

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 (OECT, 2013). While student tracking is generally less common in the U.S., it is still widely used in some areas. For example, 75 percent of U.S. schools track 8th grade mathematics students by ability (National Assessment of Education Process, 2013). However, despite the popularity of tracking, there is still debate over the merits of tracking students by ability. Of particular concern is whether it could harm students tracked into lower ability classrooms through increased exposure to lower ability peers. Unfortunately, there is little evidence on whether these differences in peer exposure induced by tracking cause subsequent differences in achievement. This is largely because many programs that assign higher-ability students to different classrooms, such as gifted and talented programs, also expose those students to different teachers and curriculum as well. 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 data from public middle schools in Thailand. Public middle schools in Thailand regularly employ ability tracking to sort students into classrooms. To measure student ability, Thai public middle schools usually have students sit for a preliminary exam before the start of 7th grade and use the score as a proxy for student ability. School officials then sort students into classrooms based on the preliminary exam score. I 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. Importantly, these cutoffs are not known to the students until after the exam is taken, making it difficult if not impossible to manipulate one's position relative to a classroom cutoff. In addition, classrooms above and below the cutoffs are required to follow the same curriculum, and have nearly identical sets of teachers. As a result, the institutional background is ideal for a regression discontinuity design, and supports the identifying assumption that all determinants of achievement other than peer quality vary smoothly across the classroom cutoff. In Section 4 I provide empirical evidence supporting this assumption. In addition, in Section 5 I show that my results are robust to the inclusion of teacher fixed effects. This is consistent with the fact that Thai teachers do not systematically choose classrooms, as well as the fact that the set of teachers is nearly identical across classroom cutoffs.

For the analysis, I implement the design using data from four public middle schools in Thailand. My data set contains preliminary score, class assignment, GPA, classroom timetable, and characteristics of 1,602 students from 10 different cutoffs. The main outcome of interest is 7th grade cumulative GPA. Importantly, GPA in Thailand is based almost entirely on student performance on multiple-choice exams for which there is no grade curving. Students in the same school also take the same exams even when they are in different classrooms and are taught by different teachers. As a result, there is very little scope for teacher bias or subjectivity to affect the GPAs in this context.

Results indicate that being tracked into classrooms with significantly higher-achieving peers does not lead to higher achievement. Specifically, I estimate that assignment to a higher ability classroom is associated with a 0.94 standard deviation increase in peer quality, as proxied by performance on the 7th grade preliminary exam. However, this exposure is associated with a statistically insignificant 0.10 standard deviation reduction in performance. Importantly, this result is robust to the bandwidth size as well as the inclusion of student characteristics and teacher fixed effects.

In addressing the effects of being tracked into classrooms with higher-achieving peers, this 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 in this context there is 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 effects of peer quality shifts due to tracking in Asia, where tracking is very common.[2] [3]

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. Similar to this study, they report no statistically significant effects, and can rule 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 also enables me to rule out effects of the magnitude found by some studies on gifted and talented programs.[4] For example, Card and Giuliano (2016) estimate that gifted and talented programs increase achievement by 0.5 standard deviations. Results here suggest

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

[2] In assessing the impact of tracking, this paper also relates to papers on the returns to school quality, as attending higher quality schools is associated with an increase in peer quality. See, for example, Hoekstra, Mouganie, and Wang, 2018; Abdulkadiroglu, Angrist, and Pathak, 2014; Clark, 2010; Dee and Lan, 2015; Dobbie and Fryer, 2014; Lucas and Mbiti, 2014; Berkowith and Hoekstra, 2011; Clark and Del Bono, 2014; Ding and Lehrer, 2007

[3] This paper also relates closely to literature on peer effects in other settings. For example, several papers estimate the effects of roommate on student outcomes (Sacerdote, 2011; Zimmerman, 2013; Carrell, Fullerton, and West, 2009).

[4] Several papers have investigated the impact of gifted and talented programs or tracked classrooms. For example, Bui, Craig, and Imberman, 2014; Card and Giuliano, 2016; Booij, Haan, and Plug, 2017; Betts, 2011

that those positive effects are likely due to factors other than the change in peer quality, such as the intensity of the curriculum and teacher quality.

These results have important implications to both parents and policy makers. Recent evidence suggests that in choosing schools, parents put much weight on peer quality (Abdulkadiroglu, Pathak, Schellenberg, and Walters, 2017). However, results have suggested that parents would be better off making decisions on factors other than peer quality, such as teacher quality or curriculum. Similarly, my results suggest that educators and policy makers should put more emphasis on other factors believed to improved student performance, and less emphasis on the role of peers. More importantly, these results have direct implications for school tracking. Specifically, findings here 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.

The rest of the paper is structured as follows. The next section explains how school system and tracking work in Thailand. Section 3 talks about the empirical strategy. Section 4 discusses the data used in this paper. Section 5 reported the results. And Section 6 concludes.

2 School System in Thailand

Thailand has a 6-3-3 school system where students go to elementary school for 6 years, middle school 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 7th grade. Compared to their U.S peers, however, Thai students have more flexibility in their choices of middle schools. Thai students are not restricted to just public middle schools in their home district. They can apply to any public middle schools they are interested in, even those that are a few hours away from their home. Many

Thai public schools also offer more than one type of curriculum. For example, in addition to the normal curriculum designated by the ministry of education, many schools also offer English-program curriculum or science-intensive curriculum as well. Students and parents can decide which school and which type of curriculum would suit them best and freely apply to the school with their preferred curriculum. The special curriculum program and the normal curriculum program are, however, typically managed completely separately from one another even when they are within the same school. This includes the admission and class sorting process as well. In this paper, I focus only on the normal curriculum program.

After choosing which schools to apply to, students can gain admission to a school through either an exam or a special quota such as a sports or donation quota. After the admission process is over, schools then typically require the newly admitted students to sit for a preliminary exam before the start of 7th grade. Apart from using the result of this preliminary exam to assess their students' readiness for the new school year, schools also use this preliminary exam score as a proxy for student ability so that they can group students into classrooms based on ability. This practice of sorting students into classrooms based on ability, or 'tracking', is a common practice in public middle schools in Thailand. For example, if there are 120 students and 3 classrooms, the 40 students who scored the highest in the preliminary exam will be grouped together in class 1, the next 40 students, i.e. those who rank 41st to 80th, will be sorted into class 2, and the 40 students who scored the lowest will be grouped together into class 3. Students who are assigned to be in the same class will stay in the same class and take courses together for the entire school year.

As mentioned earlier, students who are assigned to the same classroom will have the same timetable and therefore take the same courses from the same teachers. And since I am only focusing on classrooms in the normal curriculum program, even students in different classrooms will take the same courses and use the same textbooks as well. And while the sets of teachers assigned to different classrooms might not be completely identical, they are extremely similar. This is because every teacher teaches more than one classroom within a grade, and thus teaches both high-ability and low-ability students. As a result, the only thing different between classrooms is the level of student ability. This setting thus allows me to apply a regression discontinuity approach to estimate the impact of peer quality because comparable students who scored just above and below the threshold will be 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 enrolled in the 7th grade in four public middle schools in Bangkok between school year 2013-2014 and school year 2016-2017. The data set consists of students' preliminary exam score, class assignment, timetable, teachers assigned, GPA, and student characteristics, which include gender, height, weight, class size, birth order, and parents' marital status.

The four schools in the data set are all public middle schools from the suburban areas of Bangkok. They are not selective schools and the majority of their students are locals living nearby. For all schools, since I only look at the normal curriculum program, all classrooms are required to follow the same curriculum. I verify this by looking at classroom timetables and find 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 PE. For example, there are a few instances of students in different classrooms taking different PE courses. Some of these classrooms have basketball as PE, while the other classrooms have volleyball. This difference is likely due to the fact that schools do not have enough resources (equipment, teachers) to allow all students to take the same non-core courses. Therefore, while technically 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 the cutoffs consist of the identical courses, so that the only thing different between them is peer quality. As a result, I end up with 10 cutoffs and 1,602 students in the data set.[5]

The main outcome of interest in this paper is 7th grade cumulative GPA. Importantly, GPA in Thai middle schools is based primarily on performance on multiple-choice exams in which there are no curves. Students in the same school also take the same 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's subjectivity to affect student GPA. Since different schools in different school years could have different standards for GPA, I use standardized cumulative GPA instead of raw GPA. I standardize cumulative GPA by rescaling within each 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 students in the data set. Columns 2-4 report the same statistics, but limit the sample to only students closer to the cutoff. Specifically, for each student I calculate their distance to the cutoff using equation (1).

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

Where $prelim_i$ is the preliminary score of student i who is associated with cutoff c, $cutoff\ score_c$ is the cutoff score at cutoff c, and r_{ic} is the distance of student i's preliminary score to cutoff c. Using the distance to cutoff calculated with equation (1), columns 2, 3, and 4 show the descriptive statistics of students who are within 20 points, 10 points, and 5 points of the cutoff respectively.

Column 1 indicates that approximately 51 percent of students in my sample are female. The average height and weight are 153 cm (5 ft) and 46 kg (101 lb) respectively. Additionally,

[5]For one of the 10 cutoffs, there are 5 classrooms below the cutoff and students who score below the cutoff are randomly assigned into one of the 5 classrooms. Out of these 5 classrooms, 4 classrooms follow the exact same curriculum as the classrooms above the cutoff, while 1 classroom does not have the same non-core courses as the classrooms above the cutoff. There are 51 students in this particular classroom and I drop them from my sample. Since students are randomly assigned into this class, this should not affect results.

70 percent of students have parents that are together, which is measured independent of marital status. The average 7th grade cumulative GPA is 2.9, which translates to the average standardized cumulative GPA of roughly 0.15. Columns 2-4 show that about 96 percent of students in the sample are within 20 points from the cutoff, 65 percent are within 10 points, and 40 percent are within 5 points. And comparing to the students further away from the cutoff, those who are closer to the cutoff are more likely to be female and have a higher cumulative GPA.

4 Empirical Strategy

To disentangle the effects of peers from other confounding factors, I apply a regression discontinuity design (RDD) that compares students just above and just below the cutoff with the assumption that students just above and just below the cutoff are comparable but were as-good-as-randomly assigned into different classrooms with different types of peers.

The setting that I have in my data is, however, a little different from a standard RDD setup where there is only one cutoff. Here, I have multiple cutoffs, each with different cutoff scores. I therefore follow the method used in Pop-Eleches and Urquiola (2013) and employ the stacked RDD approach where I first normalize the cutoffs using equation (1) before pooling all normalized cutoffs together for the regression discontinuity analysis.

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

In equation (1), $prelim_i$ is student i's raw preliminary exam score and $cutoff\ score_c$ is the cutoff score of cutoff c. And normalized preliminary exam score of i at cutoff c is r_{ic}. With equation (1), all the cutoff scores are recentered to zero and each normalized preliminary score indicates the distance to cutoff instead of the raw preliminary score, which might not be comparable across cutoffs. One important thing that should be noted is that since each school could have multiple cutoffs in a school year, it is possible that some students are associated with 2 cutoffs at the same time. For example, from the example earlier where there are 3 classrooms, students who are in class 2 are associated with 2 cutoffs: the one separating class 1 and class 2, and the one separating class 2 and class 3. The students who are associated with 2 cutoffs will appear in the data set twice and have 2 different normalized preliminary scores calculated based on the two different cutoffs.

The regression discontinuity analysis in this paper then follows the standard RDD equation as shown below:

$$Y_{ic} = \alpha_1 + \beta_1 I[r_{ic} \geq 0] + \gamma_{11} r_{ic} + \gamma_{12} r_{ic} I[r_{ic} \geq 0] + \delta_c + u_{ic} \tag{2}$$

Where r_{ic} is the models' running variable, which is student i's normalized preliminary exam score. $I[r_{ic} \geq 0]$ is the indicator function of whether i is at or above the cutoff. δ_c represents a full set of cutoff dummies. And Y_{ic} is the outcome variable of student i at cutoff c.

Note that the standard errors are clustered at the student level to account for the fact that students who are associated with 2 cutoffs could appear in the data set twice. The coefficient of interest here is β_1, which indicates whether there is a discontinuity in the outcome (Y_{ic}) at the cutoff.

4.1 Validity of RD Approach

As with any RD design, the 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 assignments and therefore 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 test the validity of the assumption by providing empirical evidence consistent with the identifying assumption.

To this end, I first 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 efforts as to place themselves just above the cutoff. Moreover, there is no retake of the preliminary exam. Nevertheless, I also confirm this empirically by looking at the distribution of students' normalized preliminary scores. If students could manipulate 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, therefore, support the institutional claim that there is no cutoff manipulation.

In addition, I also test for discontinuities in observable characteristics, which include gender, weight, height, birth order, parents' marital status, and class size. First, I do so by examining each characteristic separately. Figure 2 shows graphically that all 6 characteristics are smooth across the cutoff. Regression discontinuity estimates are shown in Table 2. Results indicate that the null hypothesis that student characteristics are smooth across the threshold cannot be rejected.

Additionally, I also use student characteristics to predict student outcomes, which in this case are 7th grade cumulative GPAs, and then look at whether these predicted GPAs are smooth at the cutoff. The benefit of this method is that it allows me to attribute weight to each characteristic according to how much it contributes to student GPA. Figure 3 shows 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.

One limitation of my data is that I do not observe 7th grade cumulative GPA for roughly 14% of the students in my data set. This is because the schools did not provide me with 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 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. 7th 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 an effect of being tracked with higher-ability students.

5 Results

5.1 First-Stage Estimation

First, I examine the first-stage relationship between students' normalized preliminary exam scores and their class assignments. Specifically, I examine how crossing the 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 because there are students who received special treatment and students who opted out of the assigned classrooms.[6]

The estimated discontinuities in students' probability of being tracked into higher-ability classrooms are shown in Panel 1 of Table 3. Odd-numbered columns show estimates without any controls, while even-numbered columns show estimates with controls for student characteristics. In addition, columns 1-2 show the estimates from the regression using full sample, while column 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 Effects of Threshold Crossing on Peer Quality

Next, I turn my attention to peer quality. In this section, I examine whether crossing the cutoff and therefore having 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 classmates' standardized preliminary exam score. [7]

Panel 2 of Figure 4 shows that peer quality jumps by approximately 0.7 standard deviations when students cross the score threshold. In addition, I formally estimate this discontinuity using equation (2) explained in the empirical section with peer quality as the outcome variable. The regression discontinuity estimates are reported in Panel 2 of Table 3.

[6] In one of the schools, students who initially chose to enroll in the normal curriculum program could ask to switch to the gifted classroom after taking the preliminary exam if they could pay the higher tuition fee of the gifted classroom. I leave these 51 students in the sample in order to avoid selection bias due to their exclusion, since the decision to switch could depend on which side of the threshold they are. In Table 4A 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.

[7] 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

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

The estimates indicate that corresponding to the increase in the probability of being tracked into the higher-ability classroom, peer quality jumps by 0.7-0.8 standard deviations when students cross the cutoff. Again, my estimates are stable across bandwidth sizes and robust to the inclusion of student characteristic controls.

5.3 The Effects of Threshold Crossing on GPA

In the previous section, results indicated that crossing the cutoff is associated with an increase of approximately 0.7-0.8 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 also formally estimate the discontinuity in student GPA at the cutoff by estimating equation (2) with standardized 7th grade cumulative GPA as the outcome variable. The estimates are shown in Table 4. They are all statistically insignificant 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. The estimates are still statistically 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 biased downward. As mentioned before, institutionally, this should not be a big issue as most of the teachers in the data set teach more than one classroom and therefore ended up teaching 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 bit more negative and range from -0.11 to -0.19 (columns 3,6,9,12), but are still statistically insignificant. This suggests that if anything, students with higher ability peers may have access to higher quality teachers, causing our 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 that 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 LATE of Being Tracked into Higher-Ability Classrooms

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 shown in Table 4 divided by the increase in the likelihood of attending the higher-ability classrooms as shown in Table 3. Estimates from Panel 1 of Table 5 indicate 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 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 tracked into 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 pretty 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 is meaningful because it allows me to rule out any positive effects bigger than 0.08 standard deviations. While Figure 7 plots the LATE estimates of being tracked into high-ability classrooms which is associates 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 quite 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 peer quality could not lead to an increase in student achievement that is larger than 0.08 standard deviations.

To summarize, the results from my 2SLS estimations suggest that being tracked into better classrooms is associated with an increase of 0.94 standard deviations in peer quality, but do not lead to positive effects larger than 0.08 standard deviations on student GPA.

5.5 Subgroup analysis

One possible reason that could explain why I do not detect significant effects of peer quality is if peer effects work differently for different subgroups, e.g. male and female, and the effects for different subgroups cancel each other out. To investigate this possibility, I look at peer effects on male and female students separately. The results shown in Figure 4A and Table 3A in the Appendix section 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. Therefore, it seems highly unlikely that I do not detect effects because of this reason.

Another thing that I should note is that when students are tracked into a classroom with higher-ability peers, they automatically become small fish in a large pond and have a lower rank in the classroom. This means that my paper is similar to the school quality literature in that my reduced-form estimates include both the direct spillover effects of having high-quality peers and the effect of having lower rank in the classroom. In theory, I could potentially 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 policy-relevant reduced-form effect of being tracked into a classroom with higher ability peers.

6 Conclusion

This paper estimates the impact of being tracked into classrooms with higher-achieving peers on student achievement using data from public middle schools in Thailand. Using an RDD approach, I find that being tracked into 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 significant increases 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 Kramer (2011), who found that an increase of one standard deviation in peer quality 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.

Furthermore, my findings also enable me to speak to the literature on gifted and talented programs. Many of these papers have documented large, positive effects from enrolling in those programs. For example, Card and Giuliano (2016) estimate that gifted and talented programs increase student achievement by 0.5 standard deviations and Booij, Haan, and Plug (2017) found the effects of 0.2 standard deviations. Since my results suggest that exposure to significantly higher quality peers could not increase student achievement more than 0.08 standard deviations, it seems likely then that the large positive effects found in the gifted and talented papers are the results of other features of the programs. These features include a more intensive curriculum and better teacher quality. More generally, the results of this study suggest that parent and policymakers should perhaps focus less on peer quality when making decisions as to how to best improve educational outcomes for children.

In addition, 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

Abdulkadiroglu, A., Pathak, P.A., Schellenberg, J. and Walters, C.R., 2017. Do Parents Value School Effectiveness? (No. w23912). National Bureau of Economic Research.

Abdulkadiroglu, A., Angrist, J., Pathak, P., 2014. The Elite Illusion: Achievement effects at Boston and New York Exam Schools. Econometrica 82(1): 137-196.

Berkowitz, D., Hoekstra M., 2011. Does high school quality matter? Evidence from admis- sions data. Economics of Education Review 30 (2011) 280-288.

Betts, Julian R. 2011. ?The Economics of Tracking in Education.? In Handbook of the Economics of Education, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 341?81.

Booij, A.S., Haan, F. and Plug, E., 2017. Can Gifted and Talented Education Raise the Academic Achievement of All High-Achieving Students?.

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

Burke, M.A. and Sass, T.R., 2013. Classroom peer effects and student achievement. Journal of Labor Economics, 31(1), pp.51-82.

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

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

Clark, D., 2010. Selective Schools and Academic Achievement. B.E Journal of Economics Analysis and Policy: Advances in Economic Analysis and Policy 10(1).

Clark, D., Del Bono, E., 2014. The Long-Run Effects of Attending an Elite School: Evidence from the UK. IZA Discussion Paper No.8617 November 2014.

Dee, T., Lan, X., 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., Fryer, R.G., 2014. The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools. American Economics Journal: Applied Economics 6(3): 58-75.

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

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

NAEP., 2013, The Resurgence of Ability Grouping and Persistence of Tracking

OECD., 2013, What Makes Schools Successful? Resources, Policies and Practices- Volume IV, Ch2, Page 81

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

Lucas, A.M., Isaac M.M., 2014. Effects of School Quality on Student Achievement: Dis- continuity Evidence from Kenya. American Economic Journal: Applied Economics 6(3): 234-63.

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

Vardardottir, A., 2013. Peer effects and academic achievement: a regression discontinuity approach. Economics of Education review, 36, pp.108-121.

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

7 Figures and Tables

Figure 1: Histogram of running variable and graph of predicted cumulative GPA

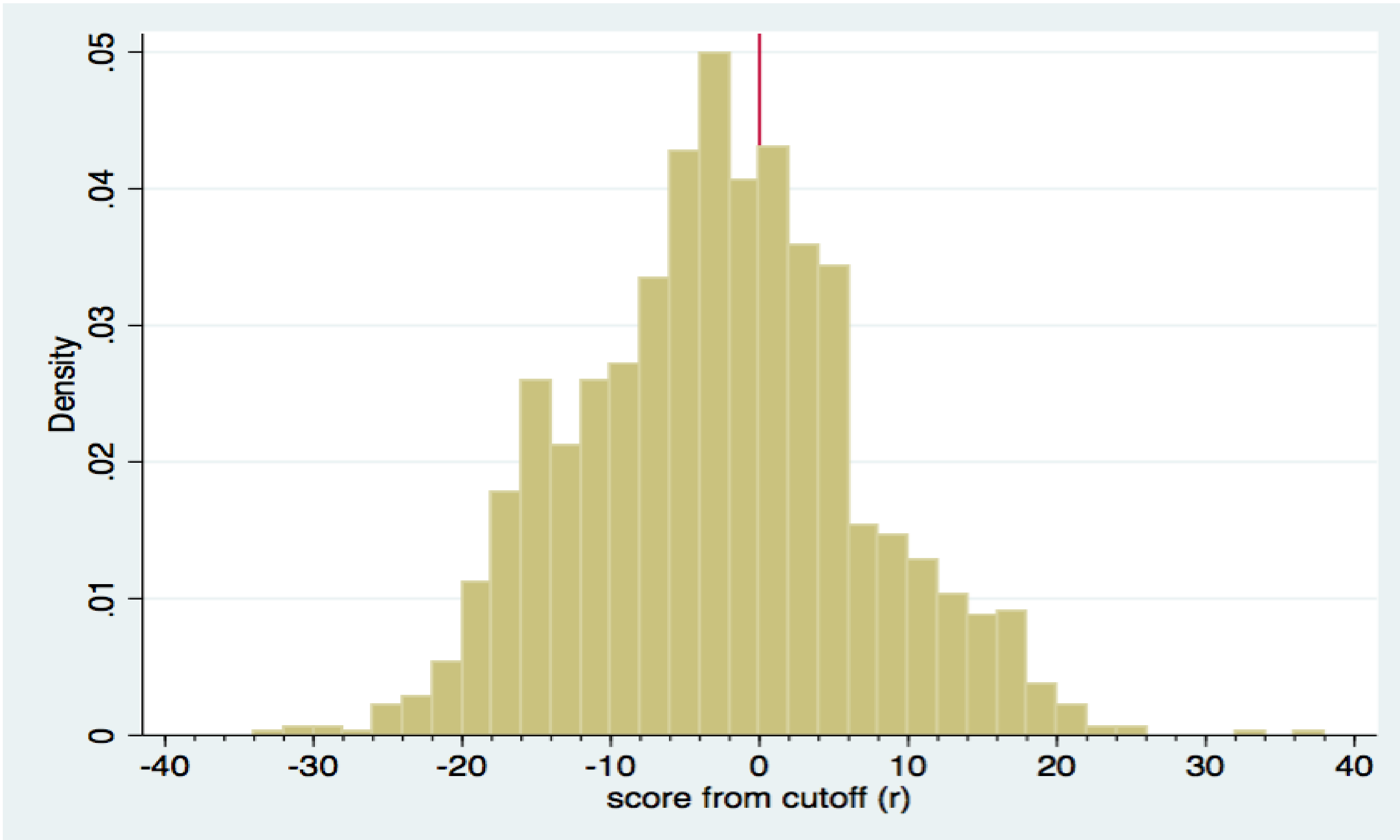


Figure 1: Note: Running variable is normalized preliminary exam score, which is the distance of student's preliminary score before 7th grade to cutoff.

Figure 2: Student characteristics across cutoff

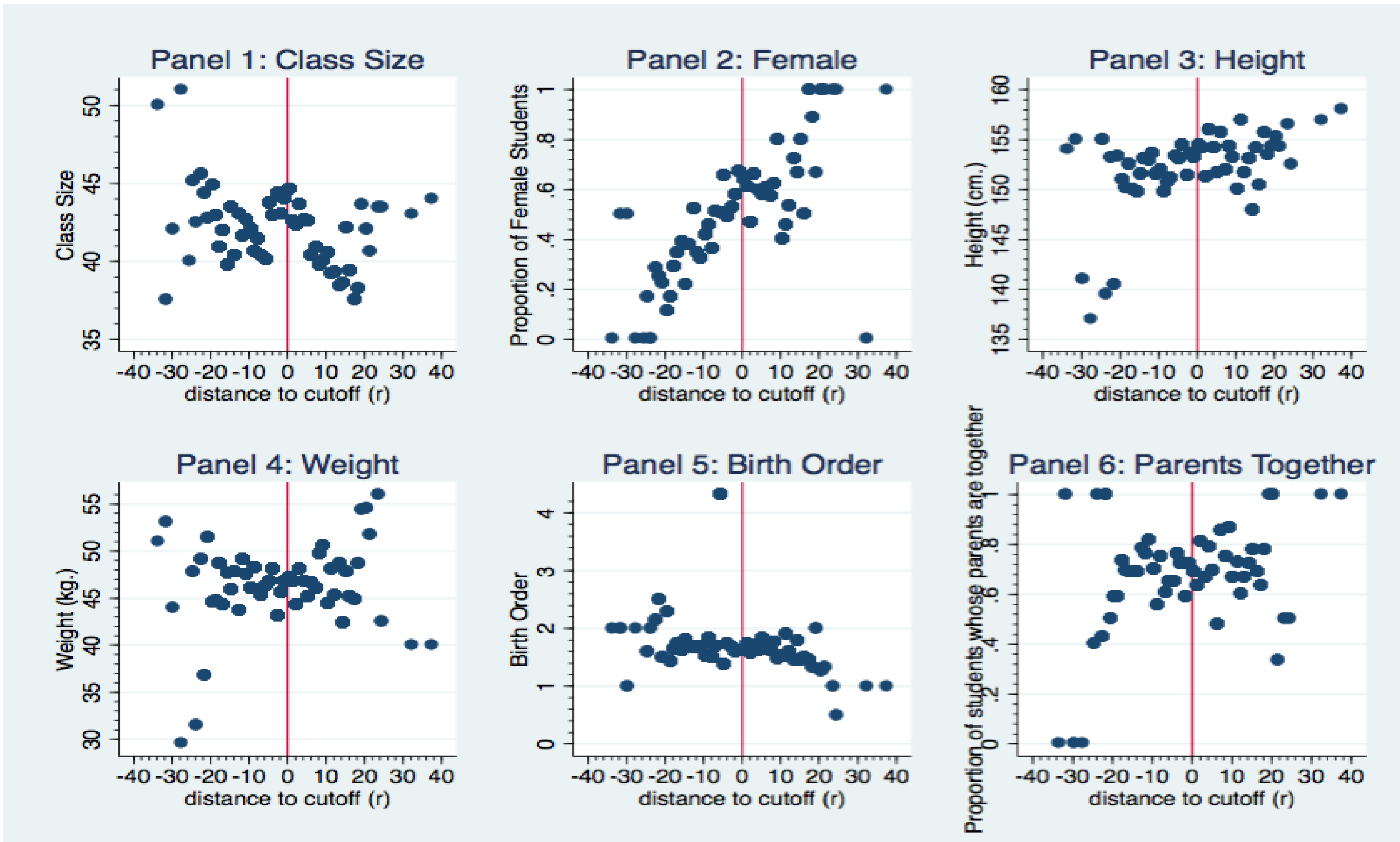
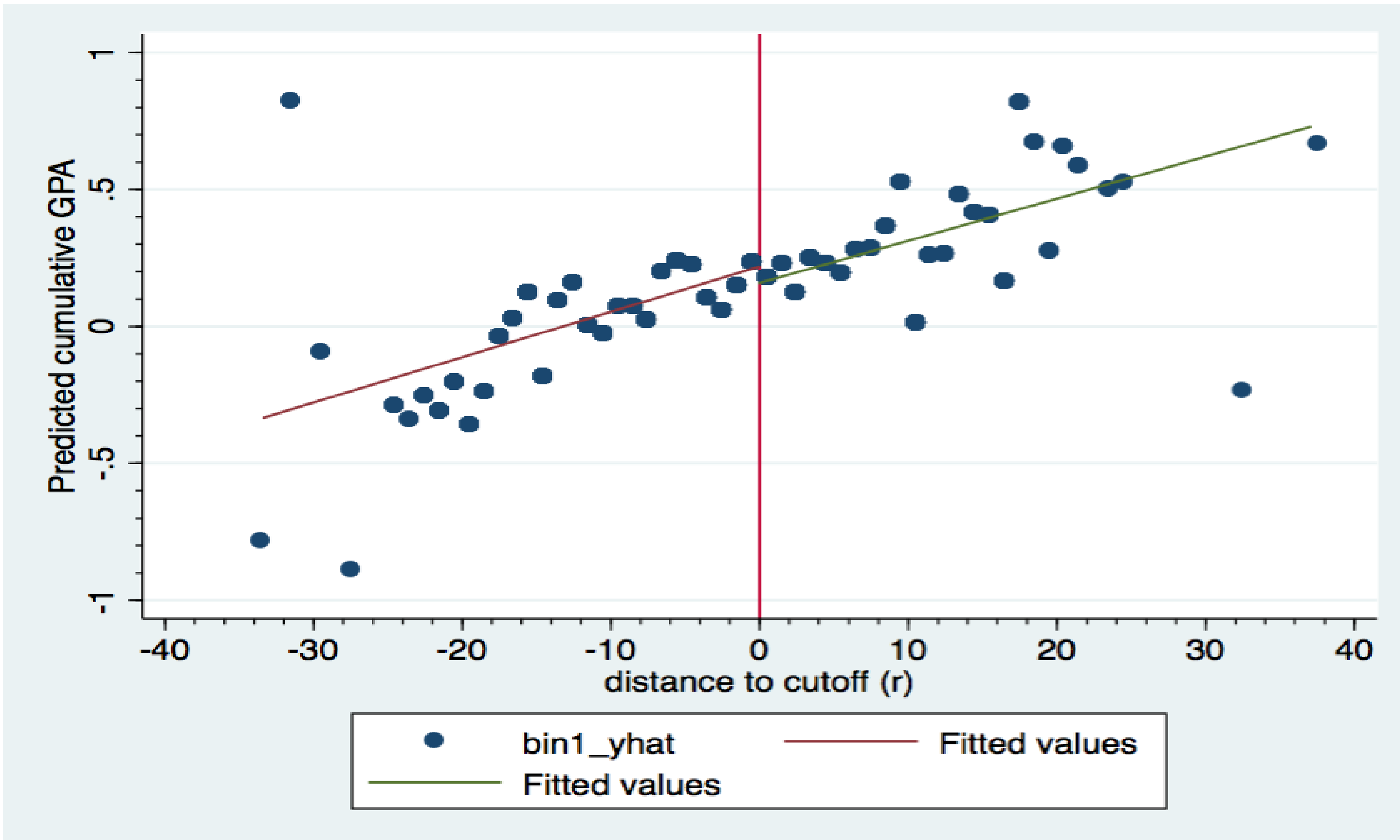
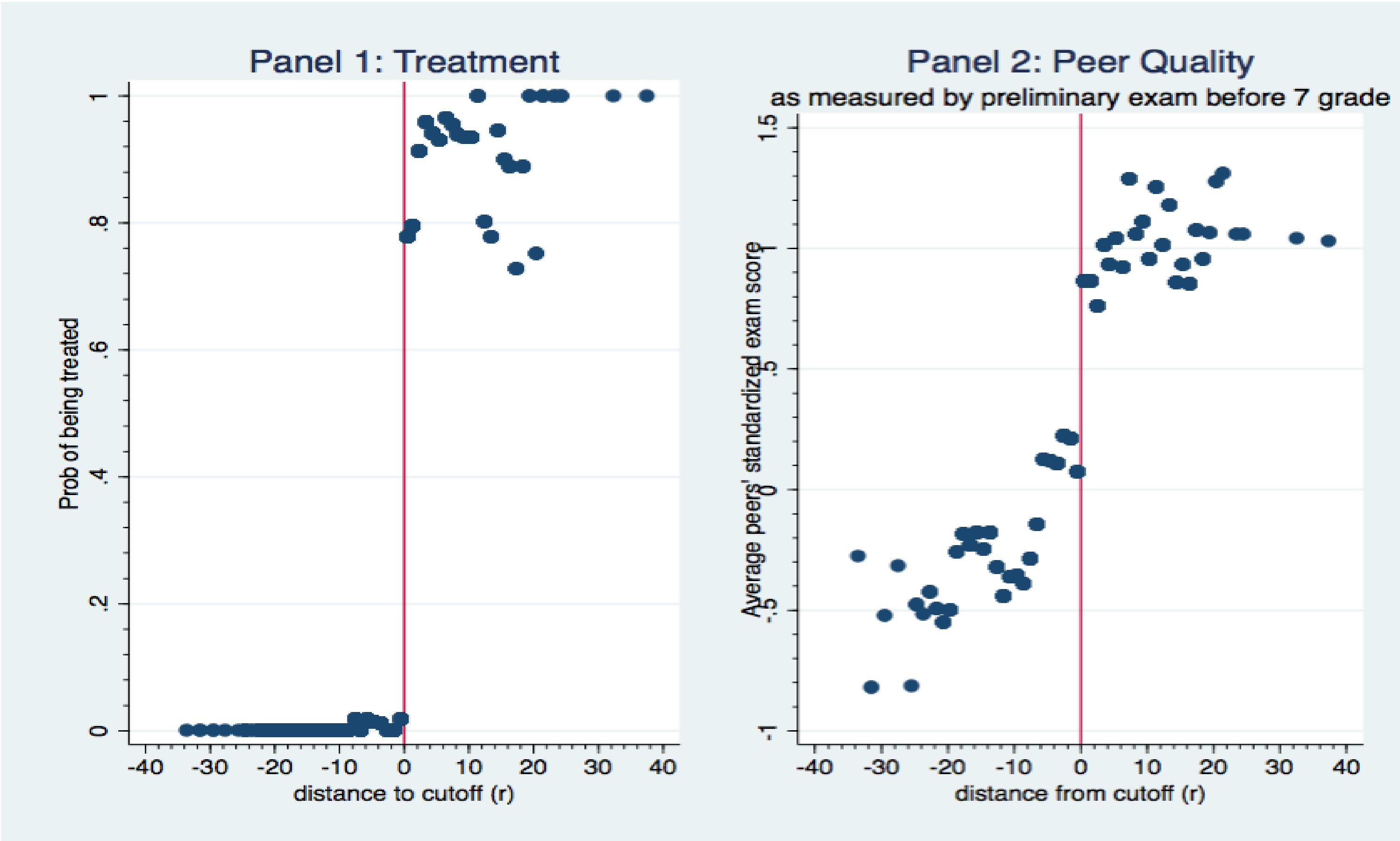


Figure 3: Predicted 7th grade cumulative GPA based on student characteristics



Predicted cumulative GPA comes from the regression of cumulative GPA on student characteristics.

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



Note: Peer quality is measured by the average of classroom peers' standardized preliminary exam score.

Figure 5: Standardized 7-th grade cumulative GPA across cutoff

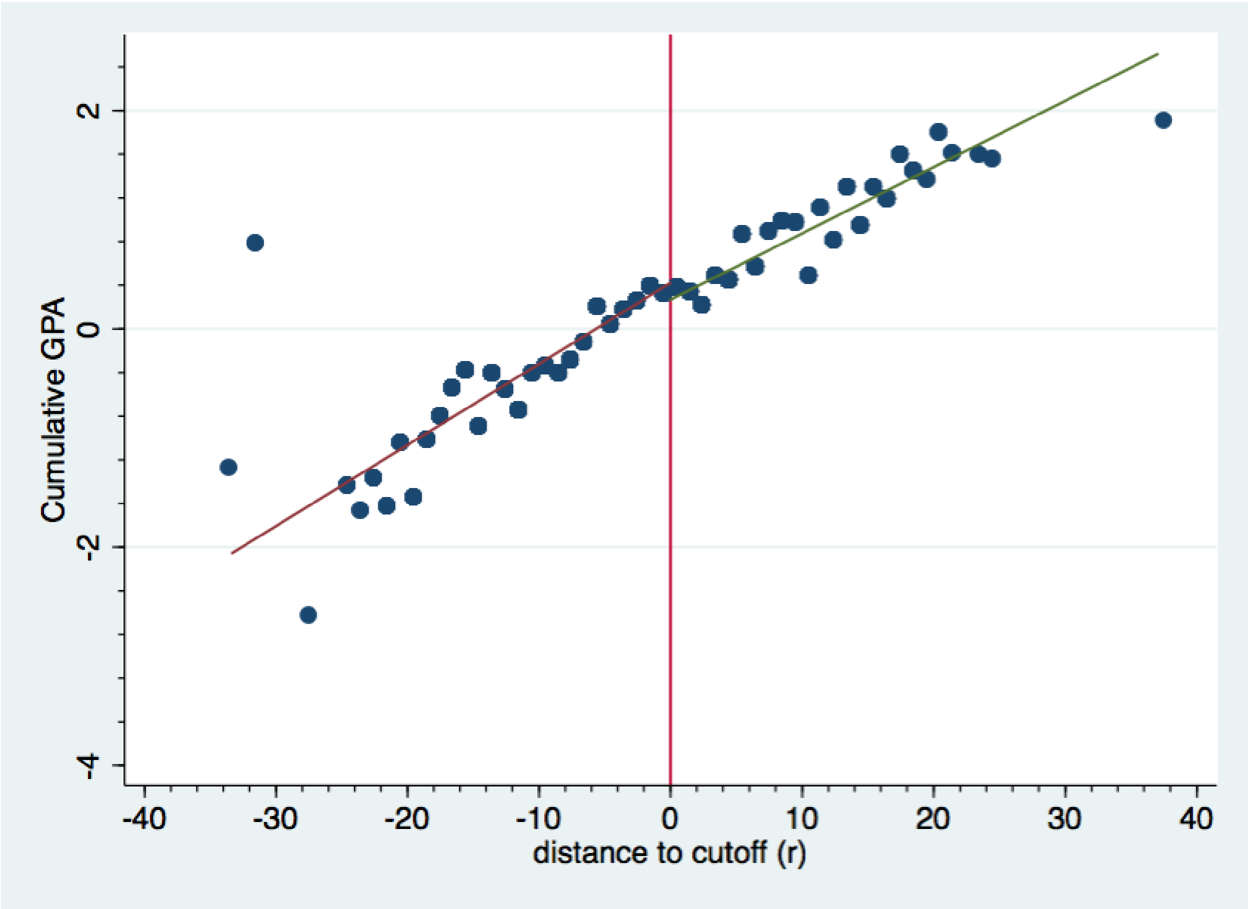
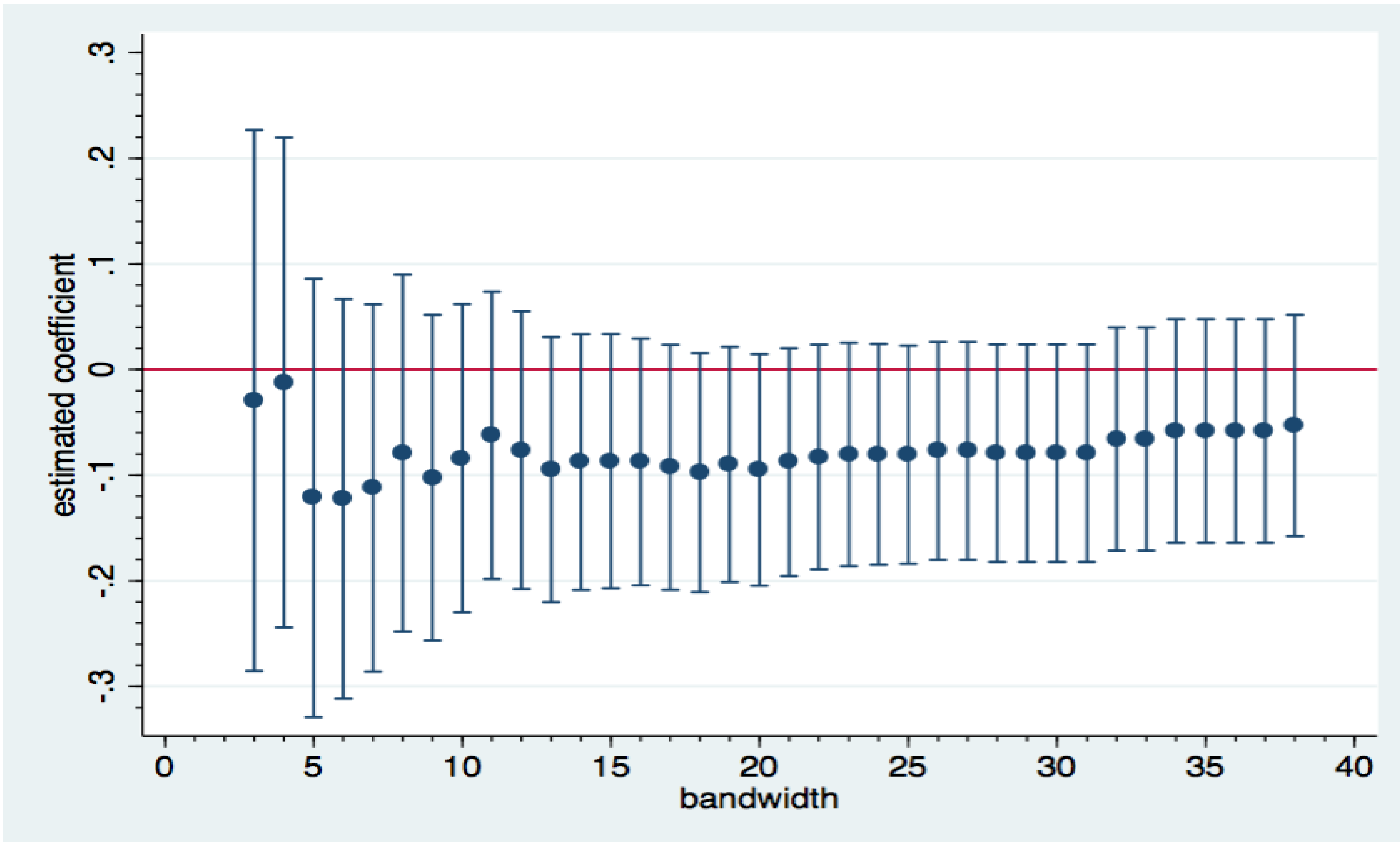


Figure 6: Reduced-form estimates using different bandwidth sizes



Note: Estimates from specification with controls for student characteristics

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

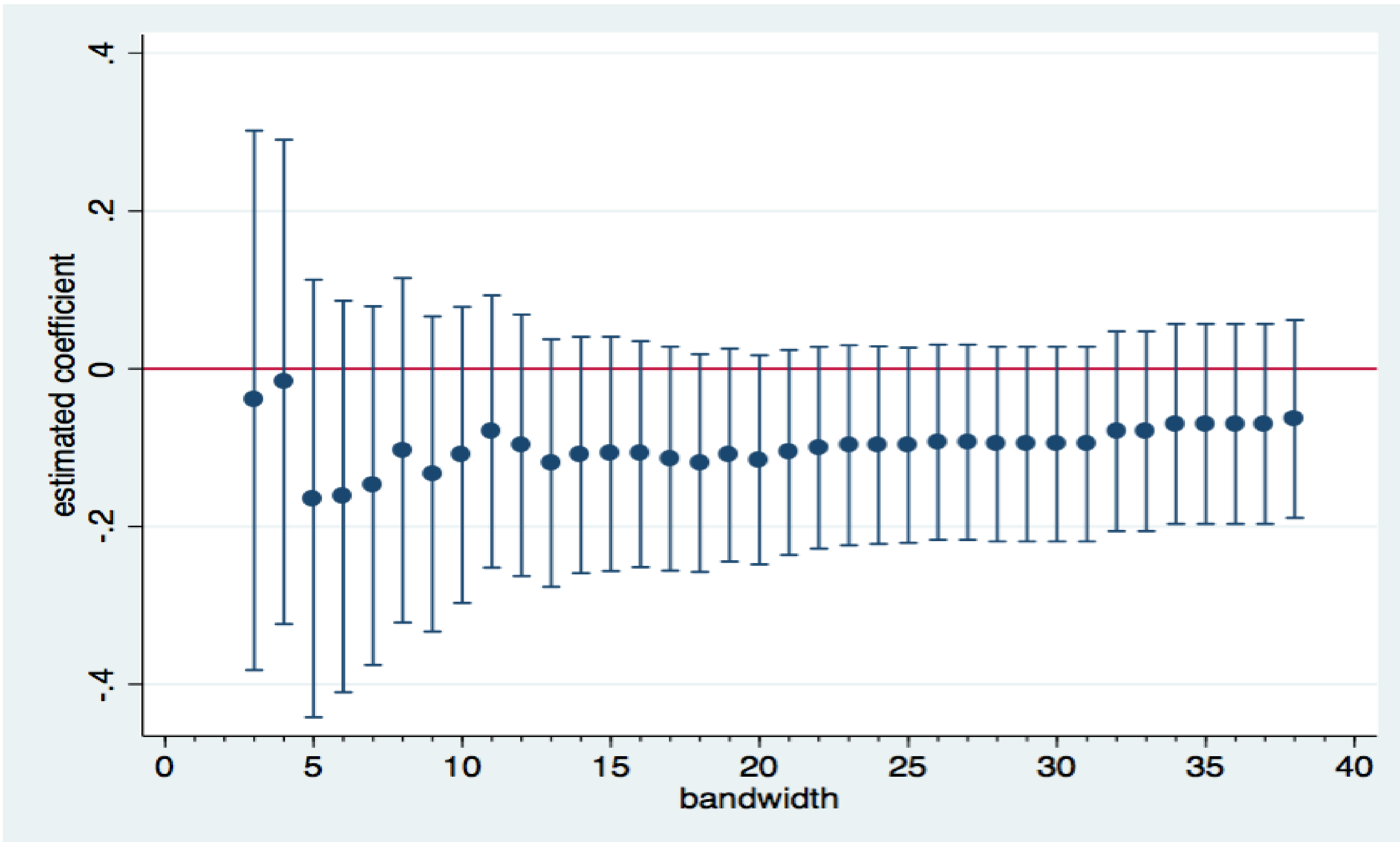


Figure 8: LATE estimates of an increase of one standard deviation in peer quality

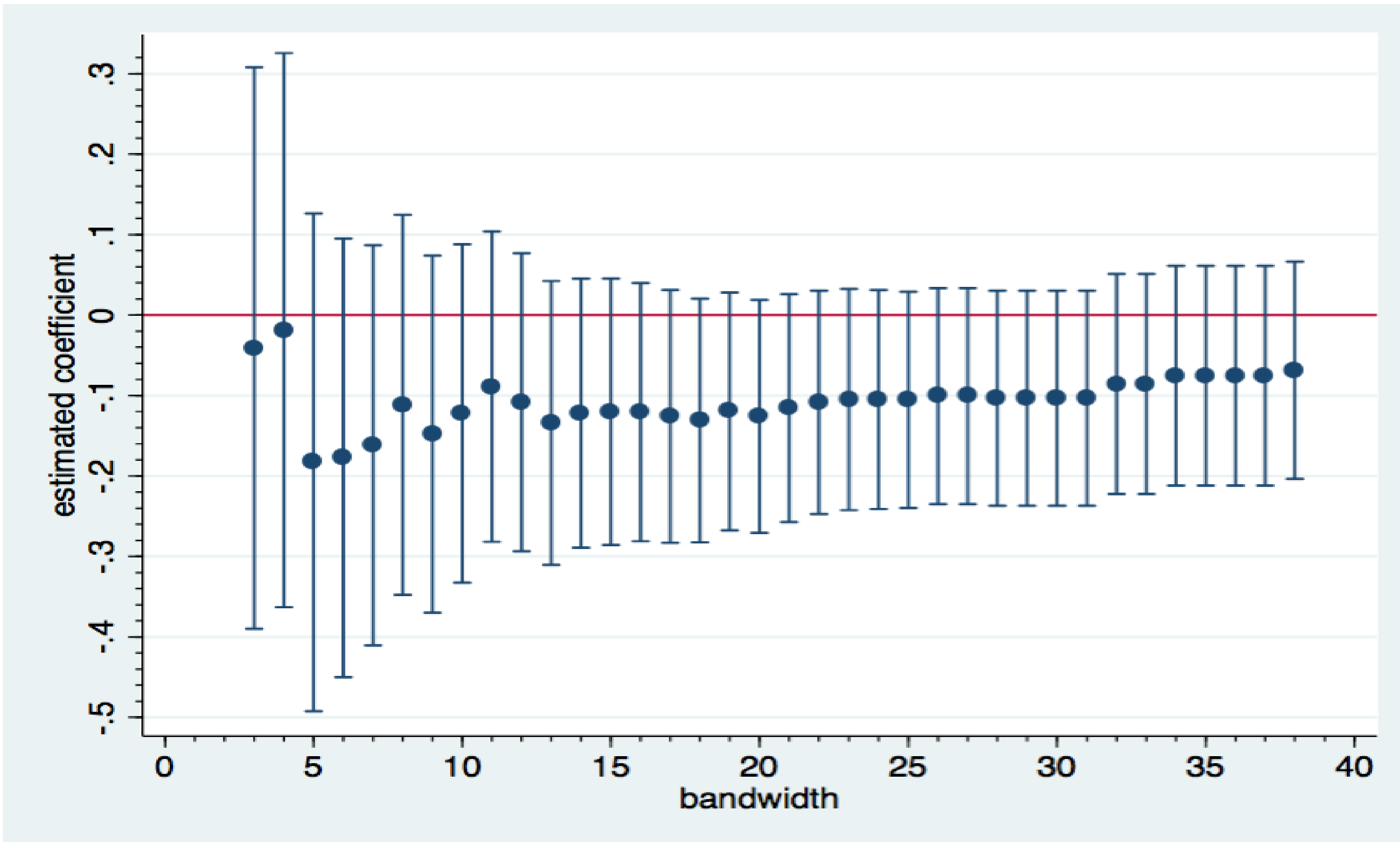


Table 1: Summary statistics

	(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
	(0.500)	(0.500)	(0.496)	(0.493)
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
	(0.460)	(0.458)	(0.460)	(0.454)
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.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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.

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

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.

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

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.

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

8 Appendix

Figure 1A: Observability of 7th-grade cumulative GPA across cutoff

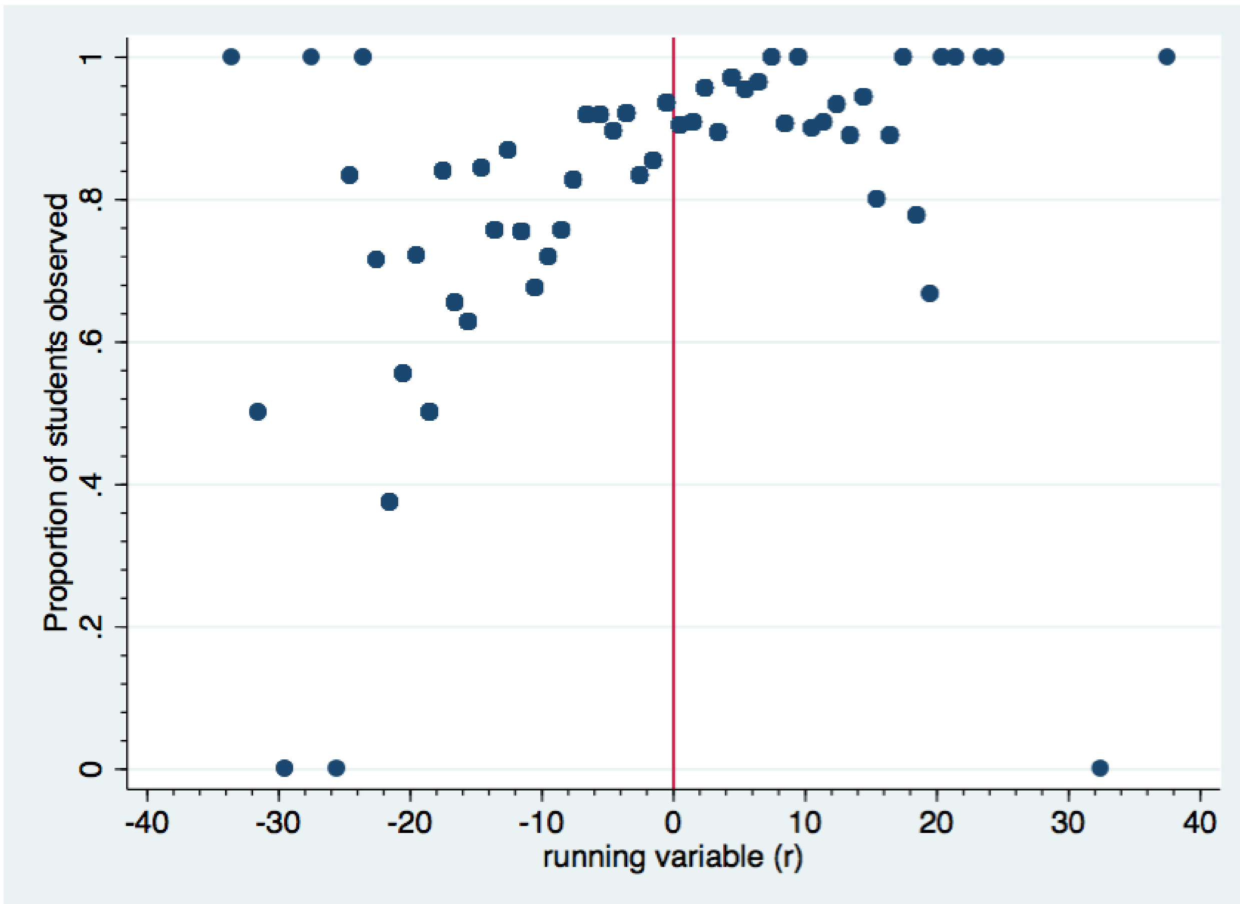


Figure 2A:Student characteristics across cutoff when only include students whose 7th-grade cumulative GPA is observed)

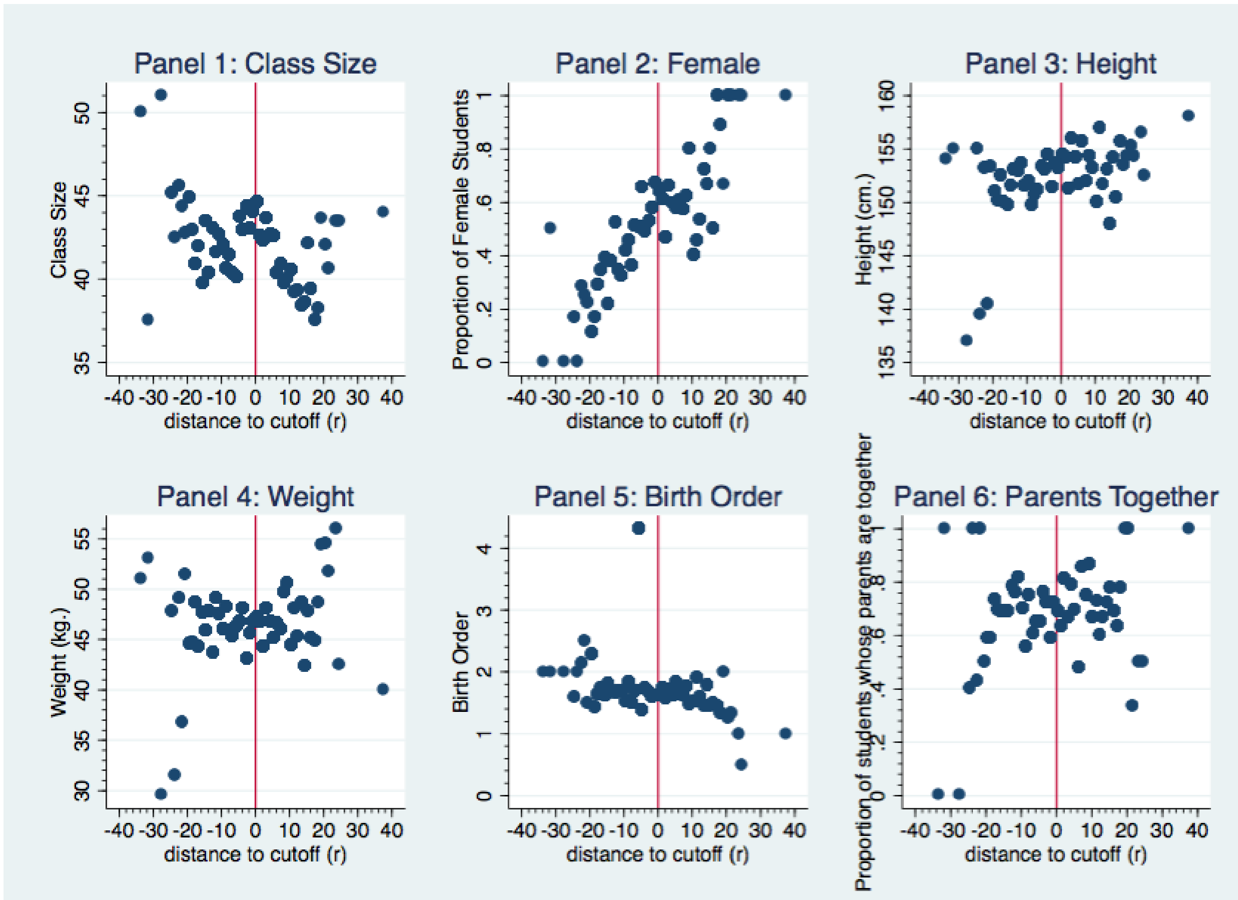


Figure 3A: Predicted 7th grade cumulative GPA based on student characteristics when only include students whose 7th-grade cumulative GPA is observed

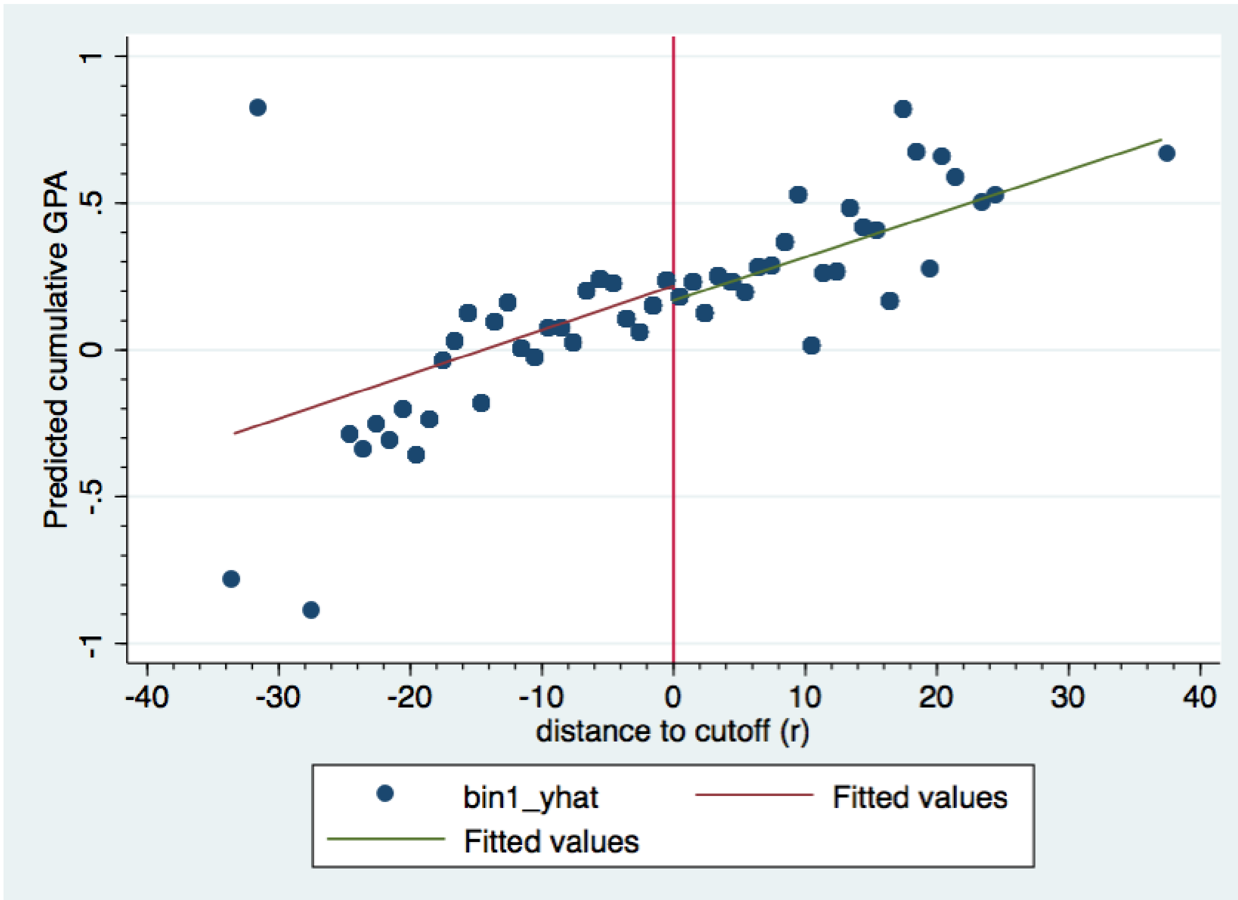


Figure 4A: 7th-grade cumulative GPA across cutoff by gender

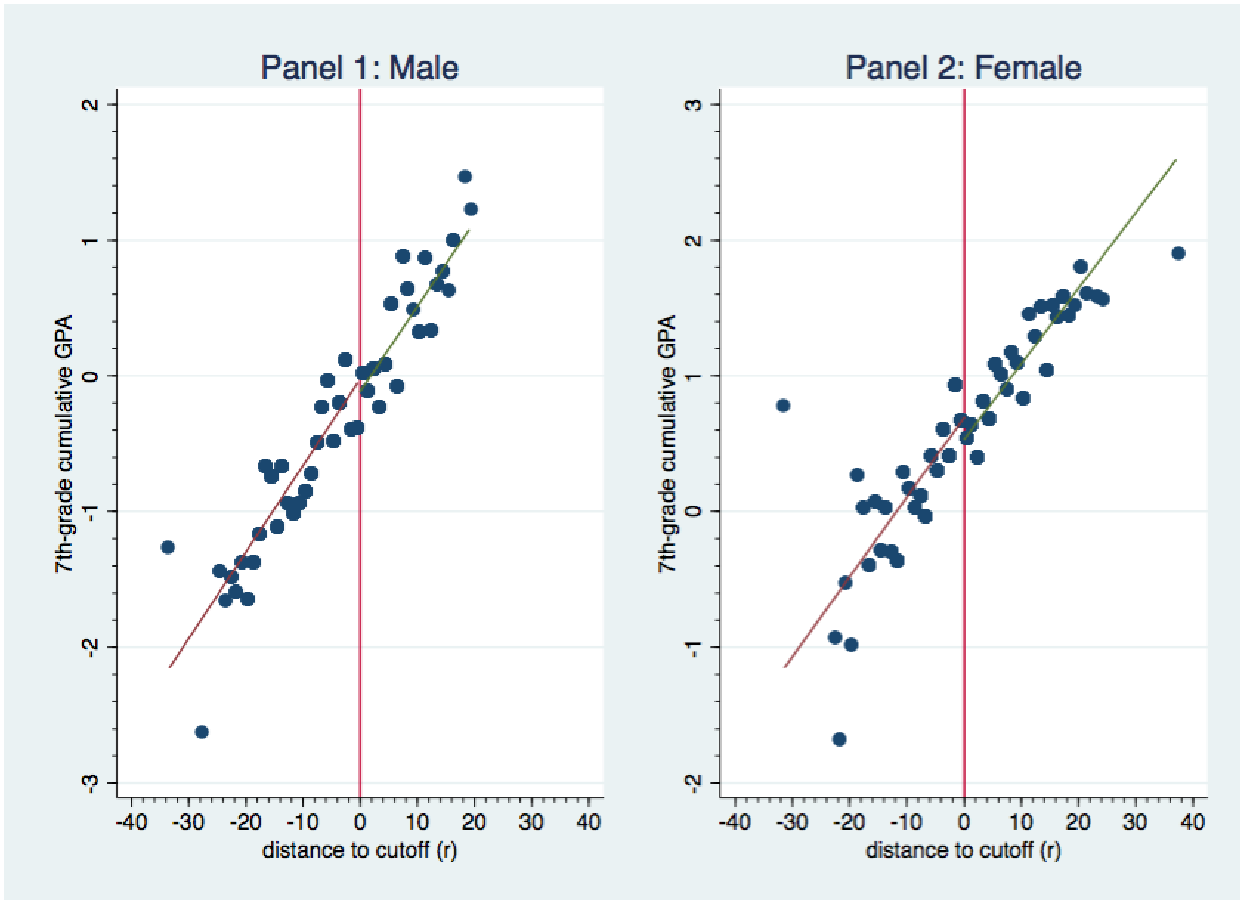


Table 1A: Regression discontinuity estimations of observability of 7th-grade cumulative GPA

	$-30 < r < 30$		$-20 < r < 20$		$-10 < r < 10$		$-5 < r < 5$	
	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.

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

Table 2A: 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	0.4230*	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.

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

Table 3A: Heterogeneous effects by gender (reduced-from estimations)

	$-30 < r < 30$		$-20 < r < 20$		$-10 < r < 10$		$-5 < r < 5$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Preliminary score above or at cutoff ($I[r \geq 0]$)	-0.0712	-0.0712	-0.0871	-0.189	-0.0832	-0.201	-0.106	-0.125
	(0.0654)	(0.0654)	(0.0692)	(0.118)	(0.0896)	(0.131)	(0.129)	(0.148)
Female & preliminary Dfem score above or at cutoff ($female \cdot I[r \geq 0]$)	-0.0136	-0.0136	-0.0130	0.0111	-0.00142	0.0183	-0.0244	0.0175
	(0.0562)	(0.0562)	(0.0567)	(0.0561)	(0.0741)	(0.0735)	(0.0946)	(0.0923)
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	1362	1362	1328	1328	947	947	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.

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

Table 4A: Reduced-form estimates

As students in one of my schools could change their mind and enroll in the gifted classroom after taking the preliminary exam, dropping students in the gifted classrooms from this school would 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.

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