American Economic Journal: Economic Policy 2014, 6(3): 30–62
http://dx.doi.org/10.1257/pol.6.3.30

Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Students[†]

By Sa A. Bui, Steven G. Craig, and Scott A. Imberman*

We evaluate the impact of Gifted and Talented (GT) programs on students through a regression discontinuity (RD) design, and by analyzing a randomized lottery for elite magnet GT schools. We show that GT students in each analysis are exposed to higher achieving peers and, in the RD sample, a more advanced curriculum. We find that achievement for marginal students neither improves nor worsens from GT services in the short run. We also find that lottery winners only perform better in science. Using a bounding analysis we cannot rule out zero, though we do not find any significant negative effects. (*JEL* H51, H75, I21, I28)

To meet the requirements of the federal No Child Left Behind Act (NCLB), school districts have had to focus on low-achieving students, potentially shifting resources away from other students including those in Gifted and Talented (GT) programs (Loveless, Farkas, and Duffett 2008; Reback 2008; Neal and Schanzenbach 2010). This is despite the popularity of GT programs, which serve over three million US students nationwide. It has been shown, for example, that GT programs help to keep students within the public school system (Figlio and Page 2002; Davis et al. 2010). What is not known, however, is whether the GT programs contribute to student learning. In this paper we provide what are, to our knowledge, the first credibly causal estimates of the effects of GT programs on student achievement.

Our work uses two unique strategies for overcoming the selection problem whereby GT students perform better than non-GT students due to factors other than program effects such as innate ability. First, we use data from a very large school district to estimate the impact of GT program enrollment on marginally eligible students using a regression discontinuity (RD) design. An important advantage of

* Bui: Department of Policy Analysis and Management, Cornell University, 105 Martha Van Rensselaer Hall, Ithaca, NY 14853-4401 (e-mail: sab427@cornell.edu); Craig: Department of Economics, University of Houston, 204 McElhinney Hall, Houston, TX 77204-5019 (e-mail: scraig@uh.edu); Imberman: Department of Economics, Michigan State University, 110 Marshall-Adams Hall, 408 W Circle Dr, East Lansing, MI 48824-1038 (e-mail: imberman@msu.edu). We would like to thank Aimee Chin, Dennis Epple, Jason Imbrogno, Chinhui Juhn, Brian Kovak, Jacob Vigdor, and seminar participants at APPAM, AEFP, NBER Economics of Education Meeting, SOLE, Carnegie Mellon University, Georgetown University, Michigan State University, University of Houston, University of Maryland, four very useful but anonymous referees, and the employees of an anonymous school district for all of their guidance and assistance in this project. Financial support from the University of Houston Small Grants program is gratefully appreciated. All errors remain our own.

[†] Go to http://dx.doi.org/10.1257/pol.6.3.30 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

our data is that while our RD approach examines marginally eligible students, due to the multidimensional nature of the admission requirements these students exhibit a relatively large dispersion of preexisting achievement levels.

Second, we separately analyze randomized admission lotteries for two premier GT magnet programs amongst students eligible for the GT program. This is similar to the analysis conducted by Cullen, Jacob, and Levitt (2006) who assess the impact of attending a better school on achievement in Chicago. Our context differs from theirs by explicitly testing the impact of a program that focuses on high ability students in a lottery context. We interpret the lottery analysis as a test of program intensity since the lottery winners study the same curriculum as losers, but are matched with stronger peers and more educated and experienced teachers.[1] This combination of approaches estimates impacts at two different portions of the student quality distribution, and evaluates two separate aspects of the GT experience. While these two research approaches do not evaluate the entirety of the effects of the GT program, they nonetheless provide the first causal evidence of program impacts.

While there is very little research on the impacts of GT education specifically, studies of related topics have found mixed results. Using standard regression techniques Argys, Rees, and Brewer (1996) find evidence that ability grouping generates positive impacts for high achieving students (and negative for low achievers), while Betts and Shkolnik (2000) find no impact using different data. Using instrumental variable techniques Figlio and Page (2002) find no significant impact of ability grouping on high achievers, but a positive impact on low achievers. On the other hand, through structural modeling Epple, Newlon, and Romano (2002) find evidence that high achievers are helped by ability grouping. Finally, a recent experiment by Duflo, Dupas, and Kremer (2011) in Kenya showed improvements from ability grouping across the achievement distribution. Particularly relevant to our study is that they find students who are near the cutoff for being tracked to the high or low group perform similarly regardless of to which group they are assigned.

In another line of research closely related to GT impacts, Abdulkadiroglu, Angrist, and Pathak (2011) and Dobbie and Fryer (2011) study the impact of attending elite high schools in Boston and New York using a regression discontinuity (RD) design. They find no impacts on marginal students. Similar studies of elite schools find positive impacts in Romania (Pop-Eleches and Urquiola 2011) and Trinidad and Tobago (Jackson 2009), as well as US higher education where Hoekstra (2009) finds significant increases in earnings from attending a flagship public university. In contrast, Clark (2010) finds little impact on achievement in England.

To our knowledge, only Bhatt (2011) and Murphy (2009) attempt to causally identify the effect of GT programs in particular on achievement.[2] Bhatt (2011) instruments for GT enrollment with the difficulty of the qualification requirements in a student's school and finds positive effects. However, such an instrument is potentially invalid if highly motivated parents of marginal students seek out schools

[1] Another possible mechanism through which GT magnets could affect students is through changes in travel time or friend networks. While we are unable to test these ideas, all students have the opportunity to attend non-GT magnet programs.

[2] See Bhatt (2011) for a review of the wider literature.

with less stringent criteria. Further, her instruments likely suffer from weak instruments bias. Using a student fixed-effects strategy, Murphy (2009) finds little math or reading improvement from being identified as GT, although these results may suffer from bias if trends in achievement are related to program entry.

There are several reasons that GT programs might help the academic performance of high achieving students. First, the peer group will be stronger on average. Second, there may be improvements in instructional resources. Third, the curriculum may be more appropriate for eligible students. Despite these advantages, it is not obvious that we should find improvements. Peer effects can be complicated because, while the mean peer improves, a student's relative position in the group also changes (Hoxby and Weingarth 2006).[3] Additionally, the interaction between educators and the peer group may be important.[4] Finally, pertinent to the lottery analysis, there is no reason to assume that the returns to program intensity are comparable to the overall returns of the GT program.

Our study utilizes a universal GT evaluation in a large urban school district in the Southwest United States (LUSD) where, since 2007, all fifth grade students have been evaluated to determine eligibility for GT services starting in sixth grade. Eligibility is determined by two well-defined cutoffs on an index score that is based on achievement tests, a nonverbal ability test, grades, teacher recommendations, and socio-economic status. We exploit these cutoffs to set up a regression discontinuity (RD) design whereby students who score just above the cutoffs are compared to those who score just below. Under certain conditions, for which we provide evidence that this analysis meets, our estimates provide the causal impact of enrolling in a GT program on achievement for students on the margin of eligibility relative to enrolling in a "regular" program. We find that achievement does not improve for students placed in GT programs. This is despite large increases in peer achievement and the likelihood of enrolling in advanced classes. We further analyze variation across schools in treatment intensity, and find no achievement gains even for students in the more intensely treated schools.

The second research strategy we employ uses randomized lotteries that determine admission to two middle schools with over-subscribed magnet GT programs. Conditional on meeting the district-wide GT eligibility requirements and completing an application, students not in the attendance zones are randomly offered admission to the district's premier magnet schools. This allows us to examine achievement differences between students who win the lottery and attend the magnet GT schools, and those who lose the lottery and attend other "neighborhood" GT programs.[5] This analysis provides evidence on the impact of extra inputs, primarily higher peer quality.

[3] For example, if peer effects are monotonic where being surrounded by higher achieving students improves one's own achievement, as found in Imberman, Kugler, and Sacerdote (2012), better peers should lead to achievement gains. For a marginal GT student, however, they are likely to go from being near the top of the regular class to being near the bottom of the GT class. Thus the students may be demoralized by reductions in their relative ranking such as proposed by Hoxby and Weingarth (2006).

[4] How the teacher targets instruction (e.g., to the median, bottom or top student) can affect the marginal student (Duflo, Dupas, and Kremer 2011). Further, Cicala, Fryer, and Spenkuch (2011) argue that the impact of moving students into an environment with higher achieving peers depends on the student's relative ranking.

[5] Students who do not attend GT magnet schools can attend GT programs in their local school, other magnet schools (based on other specializations), or charter schools. The GT programs in these other schools are called neighborhood programs, because they are not designed to attract GT students from other attendance zones.

Nonetheless, we find that the GT magnet schools provide little additional contribution to achievement over the regular GT program, with the notable exception of science.

While at first blush our findings are surprising, there are reasons to think they are plausible given the uncertainty about how students might react to more difficult course work and the structure of peer effects (Imberman, Kugler, and Sacerdote 2012). Further, they are consistent with the findings of Abdilkadiroglu, Angrist, and Pathak (2011) and Dobbie and Fryer (2011) of the impact of attending an elite magnet high school, as well as the finding from Duflo, Dupas, and Kremer (2011) of no tracking impacts on marginal students. Nonetheless, there are a few limitations to our study that should be noted. First, we can only look at outcomes one and one-half years after enrollment, so our analysis is by necessity short term. Second, while we show that the exams we use to measure achievement are capable of distinguishing between students at this upper portion of the distribution, this issue could still arise if the gifted program provides additional learning that is difficult to measure on standardized exams. Even so, given the paucity of causal estimates of the impacts of this important program in the literature, we believe that this study provides an important first step in understanding the impacts of GT programs on students.

I. The Gifted and Talented Program in LUSD

A. *Institutional Details*

LUSD is a large school district in the Southwestern United States with over 200,000 students. The district is heavily minority and low income, where the minority population is mostly Hispanic. The GT program (called Vanguard) has been in existence at least since 1972. Panel A of Table 1 provides characteristics of the GT and non-GT students we study. It shows that gifted students are less likely to be economically disadvantaged, more likely to be white, less likely to have limited English proficiency, and they perform better on achievement tests than non-GT students. To be identified as GT in LUSD, a student must meet the eligibility criteria set forth in the "gifted and talented identification matrix." The matrix for GT in 2008–2009 is provided in Figure 1. It converts scores on standardized tests, the Stanford Achievement Test for English speaking students and the Aprenda exam for Spanish speaking students with limited English proficiency, along with scores on the Naglieri Nonverbal Abilities Test (NNAT), course grades, teacher recommendations, and indicators for socio-economic status into an index we call "total matrix points."[6]

Students can meet eligibility requirements in one of two ways. The first is having 56 total matrix points, including at least 16 points from the Stanford Achievement Test or Aprenda and 10 points from the NNAT.[7] Alternatively, students can qualify

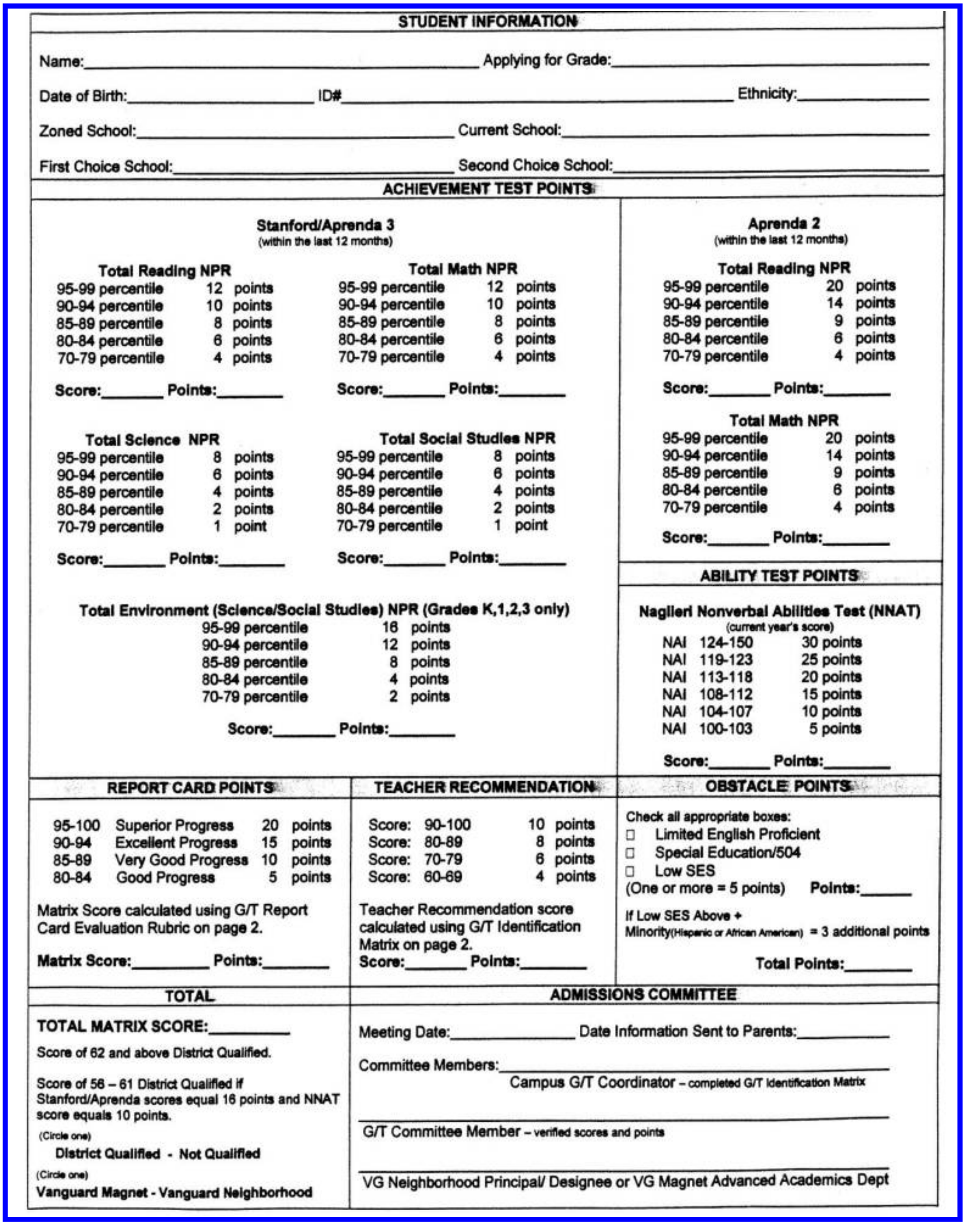
[6] For socioeconomic status, students get five extra points (out of a maximum of 108) for having limited English proficiency, being classified as special education or being classified as economically disadvantaged. Students who are members of a minority group get a further three points.

[7] Students can reach 16 points from the Stanford Tests through scores in four subjects. For example a student would qualify by being in the 90th percentile in math and the 80th percentile in reading regardless of science and social studies scores. Alternatively, a student could score in the 80th percentile in all four exams. See Figure 1 for details on the conversion of test scores to points.

Table 1—Characteristics of Students Evaluated for Middle School GT in 2007–2008

	All fifth grade students			GT magnet lottery sample		
	Gifted in 2009–2010 (seventh grade)	Not gifted in 2009–2010	Not in sample in 2009–2010	In GT magnet in 2009–2010	Not in GT magnet in 2009–2010	Not in sample in 2009–2010
Panel A. Fifth grade characteristics						
Female	0.54	0.48	0.50	0.51	0.54	0.57
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Economically disadvantaged	0.59	0.89	0.81	0.24	0.41	0.17
	(0.49)	(0.31)	(0.39)	(0.43)	(0.49)	(0.37)
LEP	0.23	0.37	0.28	0.02	0.06	0.04
	(0.42)	(0.48)	(0.45)	(0.15)	(0.24)	(0.20)
Asian	0.11	0.02	0.03	0.28	0.16	0.19
	(0.31)	(0.13)	(0.18)	(0.45)	(0.37)	(0.39)
Black	0.13	0.28	0.33	0.12	0.21	0.18
	(0.34)	(0.45)	(0.47)	(0.32)	(0.41)	(0.38)
Hispanic	0.52	0.66	0.56	0.22	0.23	0.14
	(0.50)	(0.47)	(0.50)	(0.41)	(0.42)	(0.35)
White	0.24	0.04	0.09	0.38	0.40	0.50
	(0.43)	(0.19)	(0.28)	(0.49)	(0.49)	(0.50)
Gifted	0.68	0.06	0.15	0.85	0.85	0.83
	(0.47)	(0.25)	(0.36)	(0.36)	(0.36)	(0.37)
Stanford math	1.18	–0.26	–0.02	1.61	1.39	1.72
	(0.85)	(0.83)	(1.00)	(0.79)	(0.71)	(1.03)
Stanford reading	1.10	–0.26	0.01	1.72	1.60	1.83
	(0.87)	(0.83)	(1.04)	(0.78)	(0.77)	(0.87)
Stanford language	1.07	–0.25	0.00	1.61	1.48	1.83
	(0.88)	(0.84)	(1.00)	(0.84)	(0.76)	(0.94)
Stanford social science	1.07	–0.25	0.00	1.52	1.48	1.75
	(0.86)	(0.86)	(1.01)	(0.86)	(0.84)	(0.91)
Stanford science	1.07	–0.24	0.01	1.47	1.36	1.61
	(0.88)	(0.86)	(1.00)	(0.89)	(0.79)	(0.95)
Panel B. Seventh grade outcomes						
Stanford math	1.22	–0.21	—	1.70	1.53	—
	(0.90)	(0.81)	—	(0.84)	(0.86)	—
Stanford reading	1.18	–0.21	—	1.66	1.58	—
	(0.81)	(0.82)	—	(0.66)	(0.72)	—
Stanford language	1.13	–0.19	—	1.59	1.44	—
	(0.83)	(0.84)	—	(0.80)	(0.72)	—
Stanford social science	1.16	–0.21	—	1.70	1.51	—
	(0.91)	(0.82)	—	(0.88)	(0.80)	—
Stanford science	1.16	–0.20	—	1.72	1.36	—
	(0.91)	(0.82)	—	(0.94)	(0.77)	—
Observations	1,919	8,748	3,652	291	149	102

Notes: Standard deviations in parentheses. Achievement is measured in standard deviation units within grade and year across the district. Economically disadvantaged refers to students who qualify for free lunch, reduced-price lunch, or another federal or state anti-poverty program.



STUDENT INFORMATION

Name:______ Applying for Grade:______

Date of Birth:______ ID#______ Ethnicity:______

Zoned School:______ Current School:______

First Choice School:______ Second Choice School:______

ACHIEVEMENT TEST POINTS

Stanford/Aprenda 3 (within the last 12 months)

Total Reading NPR

95-99 percentile	12 points
90-94 percentile	10 points
85-89 percentile	8 points
80-84 percentile	6 points
70-79 percentile	4 points

Score:______ Points:______

Total Math NPR

95-99 percentile	12 points
90-94 percentile	10 points
85-89 percentile	8 points
80-84 percentile	6 points
70-79 percentile	4 points

Score:______ Points:______

Total Science NPR

95-99 percentile	8 points
90-94 percentile	6 points
85-89 percentile	4 points
80-84 percentile	2 points
70-79 percentile	1 point

Score:______ Points:______

Total Social Studies NPR

95-99 percentile	8 points
90-94 percentile	6 points
85-89 percentile	4 points
80-84 percentile	2 points
70-79 percentile	1 point

Score:______ Points:______

Total Environment (Science/Social Studies) NPR (Grades K,1,2,3 only)

95-99 percentile	16 points
90-94 percentile	12 points
85-89 percentile	8 points
80-84 percentile	4 points
70-79 percentile	2 points

Score:______ Points:______

Aprenda 2 (within the last 12 months)

Total Reading NPR

95-99 percentile	20 points
90-94 percentile	14 points
85-89 percentile	9 points
80-84 percentile	6 points
70-79 percentile	4 points

Score:______ Points:______

Total Math NPR

95-99 percentile	20 points
90-94 percentile	14 points
85-89 percentile	9 points
80-84 percentile	6 points
70-79 percentile	4 points

Score:______ Points:______

ABILITY TEST POINTS

Naglieri Nonverbal Abilities Test (NNAT) (current year's score)

NAI 124-150	30 points
NAI 119-123	25 points
NAI 113-118	20 points
NAI 108-112	15 points
NAI 104-107	10 points
NAI 100-103	5 points

Score:______ Points:______

REPORT CARD POINTS

95-100	Superior Progress	20 points
90-94	Excellent Progress	15 points
85-89	Very Good Progress	10 points
80-84	Good Progress	5 points

Matrix Score calculated using G/T Report Card Evaluation Rubric on page 2.

Matrix Score:______ **Points:**______

TEACHER RECOMMENDATION

Score: 90-100	10 points
Score: 80-89	8 points
Score: 70-79	6 points
Score: 60-69	4 points

Teacher Recommendation score calculated using G/T Identification Matrix on page 2.

Score:______ **Points:**______

OBSTACLE POINTS

Check all appropriate boxes:

- ☐ Limited English Proficient
- ☐ Special Education/504
- ☐ Low SES

(One or more = 5 points) **Points:**______

If Low SES Above + Minority(Hispanic or African American) = 3 additional points

Total Points:______

TOTAL

TOTAL MATRIX SCORE:______

Score of 62 and above District Qualified.

Score of 56 – 61 District Qualified if Stanford/Aprenda scores equal 16 points and NNAT score equals 10 points.

(Circle one)
District Qualified - Not Qualified

(Circle one)
Vanguard Magnet - Vanguard Neighborhood

ADMISSIONS COMMITTEE

Meeting Date:______ Date Information Sent to Parents:______

Committee Members:______

Campus G/T Coordinator – completed G/T Identification Matrix

______

G/T Committee Member – verified scores and points

______

VG Neighborhood Principal/ Designee or VG Magnet Advanced Academics Dept

FIGURE 1. GIFTED AND TALENTED MATRIX FOR GT ENTRY IN 2008–2009

by having 62 total matrix points regardless of Stanford, Aprenda, and NNAT scores. Since 2007 all students in fifth grade are evaluated for GT, including those who participated in the GT program in elementary school.[8] These cutoffs let us causally

[8]Elementary students must re-qualify in fifth grade to maintain their classification in middle school. Students who qualify for GT in middle or high school generally keep their status through graduation, although they can be removed from GT if they perform poorly. Students can apply in later years at the school's discretion.

identify the impact of GT through a fuzzy RD design. The design is "fuzzy" because some students who qualify do not enroll in the program, while some who do not initially qualify do so later.[9]

LUSD also has two middle schools with GT magnet programs that are over-subscribed, and as a result the district uses lotteries to allocate spaces.[10] While the losers of the lottery have the opportunity to receive GT services in other schools, the magnet schools are considered to be premium schools due to their large GT populations.[11]

Table 1 also shows means from the lottery sample in panel B. The table shows that students in the lottery are significantly stronger academically than the average GT student in panel A. For example, lottery participants in panel B are shown to score about 1 standard deviation higher on standardized tests in fifth grade than the mean GT student from panel A.[12] Lottery participants are also more likely to be white, and not eligible for subsidized school lunch. Attrition bias is a potential factor, since 18.8 percent of the 542 students that enter the lottery are not in the school district by seventh grade. In fact, most of these students leave the sample before sixth grade and the leavers are different from the lottery winners. We address potential attrition bias in two ways—by reweighting the sample to look like the pre-lottery sample on observables, and through the use of a bounding analysis proposed by Engberg et al. (2010).[13]

One important aspect of interpreting the results is to understand the nature of GT programs and how they differ from regular pre-Advanced Placement (pre-AP) classes, which are taken by most marginal non-GT students. While we provide some data below with regards to peer and teacher characteristics, unfortunately we do not have data on actual differences in instruction. Nonetheless, the head of GT programs for LUSD claims that GT classes differ from non-GT by going deeper into the material, as opposed to increasing the breadth of topics. Much of this is done through the use of special projects that reflect the curriculum and cut across subjects. While teachers and principals ultimately have substantial control over what is taught, the district provides some recommended lessons. One example is a project that has students conduct independent research on an issue in health. Another has students conduct statistical analyses of everyday life events. Thus the GT program is geared towards enhancing creative and critical thinking compared to the regular program. To the extent that these skills are picked up by standardized testing, we should be able to detect whether the program enhances measured achievement.

Second, classroom composition varies between those exclusively composed of GT students, and those composed of both GT and other students who did not

[9] Another reason a student may not show up in the data as GT is if his or her school does not have enough GT certified teachers to provide the required services. This is very rare, however, as only 2 of the 41 traditional middle schools in LUSD had no GT students in seventh grade in 2009–2010.

[10] There are 8 middle schools with GT magnet programs in total (out of 41 traditional middle schools), but only 2 are over-subscribed. By seventh grade, of the 109 lottery losers that stay in LUSD, 21 enroll in one of the lottery magnet schools, only 5 attend one of the other 6 GT magnet programs, and the remainders attend a neighborhood GT program. Conversely, of the 265 lottery winners, only 3 attend one of the other 6 GT magnets in seventh grade.

[11] One of the two lottery schools also has an attendance zone. GT students from the attendance zone bypass the lottery, hence we drop any student zoned to that school from our lottery sample.

[12] Throughout this paper we standardized scale scores from each exam within grade and year across the district.

[13] We also test for selective attrition in the RD sample. While we find a small amount of differential attrition, we provide evidence below that it is not systematically related to outcomes.

qualify. Further, in some schools GT students are not enrolled in advanced (called "Vanguard") classes, but rather they receive additional instruction within the regular pre-AP class. Thus, we conduct heterogeneity analysis by Vanguard class enrollment and the percent GT in a course to test for these differences.[14]

Finally, GT teachers are required to have acquired certain credentials. In particular they must take 30 hours of GT professional development and receive training in the district's GT curriculum. In practice, however, we show below that there is little difference in GT certification rates between teachers of GT enrollees and nonenrollees as certified teachers also teach regular classes. Nonetheless, we do find that GT teachers are slightly more experienced than non-GT teachers.

B. *Data for the Regression Discontinuity Analysis*

Our data consists of administrative records from 2007–2008 to 2009–2010. While we have data for universal assessments conducted in 2006–2007, many schools were given exemptions from the new rules that year in order to allow for an orderly transition to the new system. As such, we start our sample with fifth graders in 2007–2008, the second year of the mandatory GT assessment, and examine outcomes in seventh grade during the 2009–2010 school year. For outcomes we use scale scores standardized across LUSD within grade and year on the Stanford Achievement Test. The test results are in standard deviation units relative to the district-wide distribution in a grade for math, reading, language, science and social studies exams.[15] After restricting the sample to a 15 unit band around the cutoff, we look at achievement of 4,055 students in the seventh grade cohort who were evaluated for GT in fifth grade. Of those students, 1,509 are eligible for GT, and 2,546 are not.

C. *Data for the Lottery Analysis*

Our lottery sample is derived from the set of fifth grade students identified as GT in 2007–2008 who apply for admission to one of the two middle schools with an over-subscribed GT magnet program.[16] We restrict our analysis to students who are enrolled in LUSD in fifth grade as these are the only students for whom we have pre-lottery characteristics. Also, this restriction reduces the likelihood of endogenous attrition as students who enter the lottery from outside LUSD are more likely to leave if they lose the lottery. In addition, we drop students zoned to the regular program for the one lottery school with an attendance zone.

While admission for nonzoned students is determined by lottery, our data does not directly provide the lottery ranking or outcome. Instead we identify whether a student is offered admission including those initially on a wait list.[17] In total the

[14] Since we do not have section identifiers we cannot place students in a specific class nor identify the exact classroom peers. Thus we use teacher-grade-course cells to proxy for classes. This is described in more detail below.

[15] Less than 0.1 percent of seventh grade and 0.5 percent of fifth grade students only take a Spanish-language alternative to the Stanford Achievement Tests. We drop these students.

[16] The application process involves a single form where students rank to up to three of the eight magnet schools. LUSD prohibits students from receiving offers from more than one of the GT lottery magnet schools.

[17] One caveat to the lottery is that GT eligible students with an older sibling in the school are exempted from the lottery, and automatically admitted. Unfortunately, LUSD was unable to provide data on siblings, but we believe

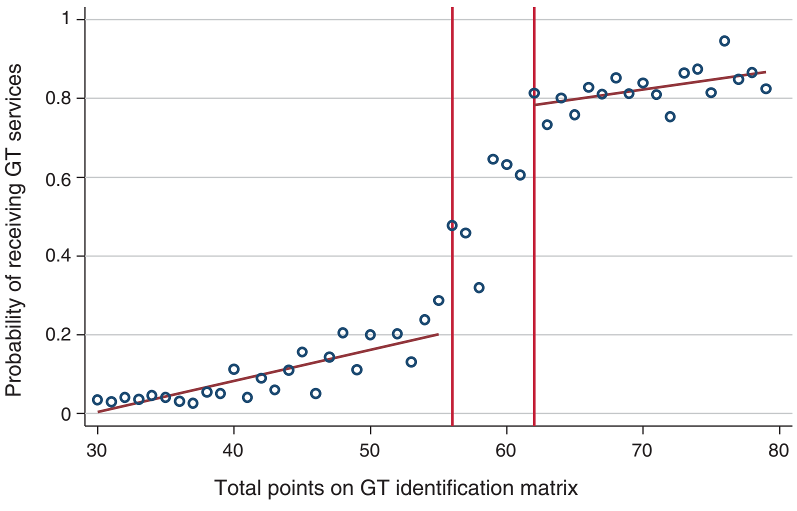


Figure 2. Gifted Status in the Seventh Grade by Fifth Grade Matrix Score

sample includes 542 students who participate in a lottery. Of these 394 are offered admission and 148 are not. By seventh grade 440 students including 331 winners (84 percent) and 109 losers (74 percent) remain in LUSD. The treatment received by the lottery losers varies, as they can attend GT classes in their neighborhood school, a charter, or a non-GT magnet school. Since there is some noncompliance with the lotteries we employ a 2SLS strategy that instruments GT magnet attendance with lottery outcomes.[18]

II. Models and Specifications

A. *GT Program Evaluation Using Regression Discontinuity*

The goal of the RD analysis is to estimate the local average treatment effect of providing gifted services to students who are on the margin of GT qualification. Figure 2 shows GT identification two years after evaluation (seventh grade) as a function of the students' matrix points. The gradual increase up to 28 percent at the first cutoff (of students with a matrix score of 56) reflects missing matrix components, qualifying in seventh grade and the district's appeal process. Upon reaching the first threshold, GT enrollment jumps to 45 percent. Enrollment increases further at a steep rate between the two cutoffs, hitting 79 percent at the second cutoff (62 matrix points). After reaching the second cutoff, GT enrollment slightly increases further to 82 percent.

siblings have a negligible impact on our estimates. First, siblings still need to apply and qualify for GT. Second, our lottery sample is very well balanced on observables between "winners," including those accepted under the sibling rule, and "losers," thus indicating selection effects are unlikely. Finally, even if older siblings potentially offer advantages, lottery losers may have older siblings in the school they ultimately attend.

[18] By seventh grade 67 percent of lottery winners attend a magnet with a lottery while 17 percent attend another school and 16 percent leave the district. For lottery losers, 18 percent attend a lottery campus in seventh grade while 56 percent attend a different school and 26 percent leave the district.

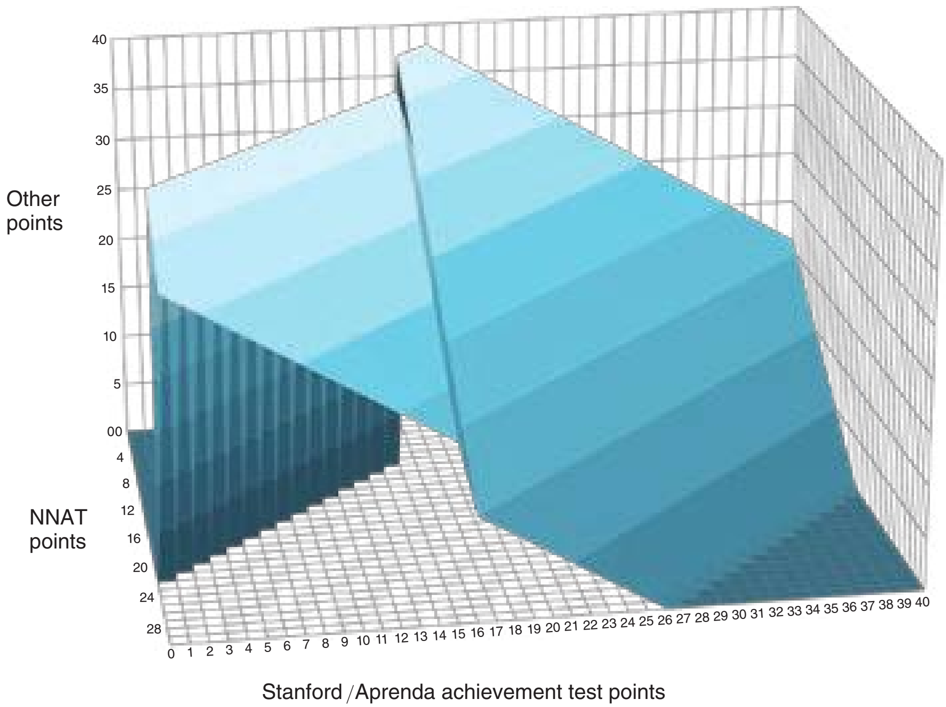


Figure 3. Surface Plot of GT Qualification by Matrix Points

Given that the increase in GT over this range is steep but not discontinuous, we convert the two thresholds into a single cutoff. To do this we map components of the matrix scores—Stanford/Aprenda points, NNAT points and other points (for socio-economic status, LEP, disability, grades, and teacher recommendations)—into three-dimensional space.[19] We then take the Euclidean distance from each student's total matrix points to the closest integer combination on the surface.[20] The resulting value, the distance to the qualification threshold ("distance"), equals zero if the student just barely qualifies for GT. We show this graphically in Figure 3. Each axis reflects one of the three portions of the matrix score that determines eligibility—NNAT points, Stanford/Aprenda points, and other points, which includes socio-economic status, grades, and teacher recommendations. Students who are on or above the surface generated by this mapping are eligible for GT while those below or behind it are ineligible.

Figure 4 shows GT enrollment in seventh grade as a function of Euclidean distance from the threshold. Students just below the cutoff have a 25 percent likelihood of being in GT while students just above have a likelihood of 79 percent.[21] The

[19] We also estimate models using each of the two cutoffs individually for subsamples affected by each cutoff and find similar results.

[20] The Euclidean distance is measured as $Dist_i = \sqrt{(Stanf_i - Stanf_s)^2 + (NNAT_i - NNAT_s)^2 + (Other_i - Other_s)^2}$, where i refers to the student's own score and s refers to the closest integer combination on the surface. Online Appendix Figure 1 shows that our Euclidean distance measure correlates very well with total matrix points. We thank Jake Vigdor for first suggesting this method to us.

[21] Note that, by construction, "distance" has an empty mass between 0 and 1, and between −1 and 0, since the smallest distance to another integer point is 1.

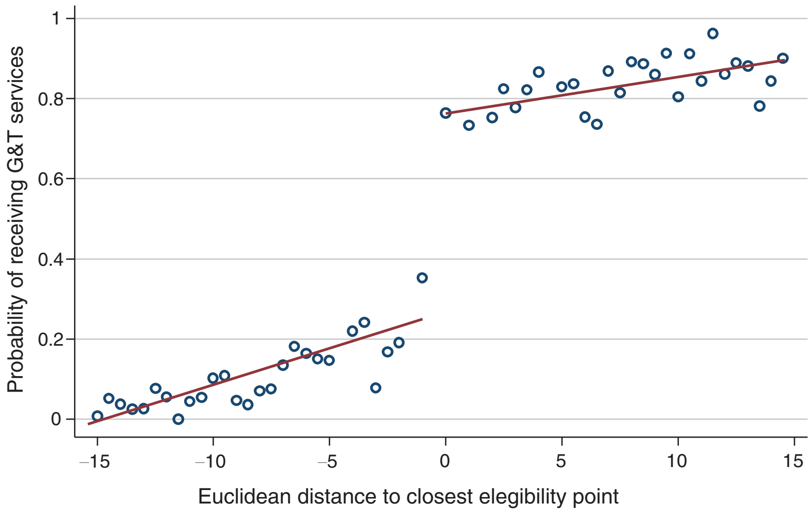


Figure 4. Gifted Status in Seventh Grade by Distance to Boundary Based on Fifth Grade Matrix Points

fuzzy RD model is based on a two-stage least squares regression within a range of values that includes the cutoff (Hahn, Todd, and Van der Klaauw 2001; Lee and Lemieux 2010). We generally use fifteen distance units below and above the cutoff for our bandwidth since the relationship between distance and achievement is close to linear over this range, allowing us to use a linear smoother.[22] Hence, we estimate the following two-stage least squares (2SLS) model:

$$(1)\quad GT_{i,t+k} = \delta + \gamma Above_{it} + \rho_1 Dist_{it} + \rho_2 Dist_{it} \times Above_{it} + \mathbf{\Omega X}_{it} + \mu_{ijt+k}$$

$$(2)\quad Y_{i,t+k} = \alpha + \beta \widehat{GT}_{i,t+k} + \lambda_1 Dist_{it} + \lambda_2 Dist_{it} \times Above_{it} + \mathbf{\Phi X}_{it} + \varepsilon_{ijt+k},$$

where $Above_{it}$ is an indicator for whether student i in year t has a distance measure at or above the cutoff, *Dist* is the Euclidean distance of the student's matrix score to the eligibility cutoff, and $\mathbf{X}$ is a set of preexisting (fifth grade) observable characteristics which includes the fifth grade dependent variable (e.g., lagged achievement), gender, ethnicity, gifted status, economic disadvantaged status, and LEP status. *GT* is an indicator for whether the student is enrolled in a GT program in year $t + k$ and Y is a test score (in standard deviation units) in that year.[23] Since students are tested in January of each year, we focus on outcomes in the second year after evaluation (seventh grade) as assessment in the first year will only provide five months of program exposure.[24]

A particularly important concern with this model is that it essentially provides a "reduced-form" assessment of the marginal impact of GT services. Since there are multiple and varying underlying treatments relative to the regular program that go

[22] We show later that our results are not sensitive to the choice of bandwidth.

[23] We also examine attendance and discipline as outputs. Not surprisingly given the high performance of this student group, neither is affected by GT participation. These results are available by request.

[24] We also analyze sixth grade outcomes, and the results are qualitatively similar. Further, we also find our measures of treatment are consistent between sixth and seventh grade. These results are available by request.

into a particular student's GT experience, this model will provide a weighted average of the impact of all those treatment effects. To provide insight into the nature of these treatment effects, we provide a series of analyses that identify the different treatments included in the GT program. We estimate heterogeneous effects by aspects of specific programs at each school, and we exploit cross-school variation in the changes in treatments to identify impacts for students who experience different changes in underlying treatments. Combined, these analyses provide a broad view of the impacts of GT and the elements that go into a GT program.

B. *GT Magnet Evaluation Using School Lotteries*

Our second analysis compares the performance of students who win a lottery to attend one of the premier magnet GT programs to those who lose the lottery and attend either a neighborhood GT program in the district, a magnet school based on a different specialty, or a charter school. This analysis allows us to examine a different segment of the GT student quality distribution. Since both winners and losers receive GT services, however, this model does not evaluate the impact of GT treatment per say, but rather it estimates the effect of providing a more intense treatment. Hence, in the lottery sample we estimate the following 2SLS model conditional on applying for admission to a magnet program with a lottery:

$$GTMagnet_{ijt+k} = \delta + \gamma Admitted_{ijt} + \mathbf{\Omega X}_{it} + \nu_j + \mu_{ijt} \tag{3}$$

$$Y_{ijt+k} = \alpha + \beta GT\widehat{Magnet}_{ijt+k} + \mathbf{\Phi X}_{it} + \eta_j + \epsilon_{ijt}, \tag{4}$$

where *GTMagnet* is an indicator for attending any GT magnet program, including those that do not hold a lottery, *Admitted* is an indicator for being offered a slot at a program with a lottery, and **X** is a set of student level controls.[25] Finally, since each school holds separate lotteries we include ν_j and η_j in the model as lottery fixed-effects.[26]

Understanding the difference between treatment and comparison groups is somewhat more complicated in the lottery analysis than in the RD since students that lose the lottery may attend any other magnet school to which they gain admission. Further, all but two middle schools have a GT program. Thus we interpret the lottery analysis as a test of program intensity in the two premier magnet programs compared to a weighted average of the district's nonmagnet GT programs.[27] Since the actual curriculum is similar in both local and magnet GT programs, we show below that this treatment mainly involves an increase in peer achievement, percent of peers who are GT, and observable teacher quality.

[25] Our results are similar if we define *GTMagnet* as being equal to one only if the student attends one of the lottery schools, as there are only a few students who attend nonlottery GT magnet schools.

[26] Since we focus only on one cohort, fifth graders in 2007–2008 (who are in seventh grade in 2009–2010), there is a single lottery fixed-effect indicator in each regression.

[27] While we set *GTMagnet* equal to one if the student attends a nonlottery magnet, the local average treatment effect is identified off of students induced to attend a magnet due to winning the lottery, which will usually be a lottery magnet.

C. *Assessment of Achievement*

We use Stanford Achievement Test (SAT) scores as our achievement measure. While it is unlikely that multiple choice exams are able to measure all of the potential educational gains from GT programs, the SAT exam has the advantage of being easily quantifiable, comparable across students from different environments, and proximate in time to the program.[28] More important perhaps is that achievement tests are frequently used as a partial measure for identifying gifted students, not only by this school district but in the academic community at large.[29] Further, we stress that, despite the limitations of this outcome, we are not aware of any other study that has been able to credibly establish the causal impacts of GT programs on any student outcome.

Another possible issue with the use of SAT is the existence of "ceiling effects," whereby if students consistently achieve the maximum score their true achievement is censored from above. We believe that the potential bias from GT students being more likely to reach the ceiling than others is negligible, as even lottery GT students have substantial room for further gains. This is exhibited in the distributions of raw scores shown in online Appendix Figures 2–6 for the RD sample and 7–11 for the lottery sample. In the RD sample, there are very few students who score near the maximum and the mean is between 58 percent and 77 percent of the maximum depending on the exam. Even the stronger lottery students, however, score below 90 percent in all subjects, and as low as 75 percent. Thus if achievement gains occur in the GT program, they should be observable in the Stanford tests.

III. Regression Discontinuity Estimates of the Impact of GT on Achievement

A. *Tests of Validity of RD Design*

In Figure 5 we provide density plots around the eligibility discontinuity. The figure shows that the variation in densities across the discontinuity are similar in size to changes at other parts of the distribution, suggesting that bias in our results from manipulation is unlikely (Lee and Lemieux 2010).[30] In Table 2, we provide tests of discontinuities in preexisting (fifth grade) student characteristics.[31] We find no discontinuities in columns 1–14 with the exception of math achievement. Given that math is

[28] Course grades are conflated by the differing difficulty levels of courses and the fact that grades are subjective and a substantial portion are scaled to be relative to other students in the same course. In results not shown here (available by request) we find that average course grades actually fall for treated students in both the RD and lottery analysis. Disciplinary infractions also are of little use as the incidence of these amongst the population studied is very small.

[29] See for example Naglieri and Ronning (2000).

[30] Ideally one would like to conduct McCrary's (2008) test. Since there are no observations between 1 and 0 or -1 and 0 (see note 21) and there is positive mass between integers further out, this could mistakenly generate a positive result. Hence, instead we test for discontinuities at the two cutoffs in the total matrix points distribution to check for possible manipulation. In both cases the test is statistically insignificant. We also provide graphical evidence on the distribution of matrix points in online Appendix Figure 12.

[31] By construction some heaping of the data may occur in the transformation from matrix scores to Euclidean distances, a potential source of bias (Barreca, Lindo, and Waddell 2011). We do not, however, find evidence of heaping in matrix scores. Our bandwidths are wide enough to include substantial observations both at heaping and nonheaping points on both sides of the cutoff. Further, we show later that our results are quite robust to choice of bandwidth.

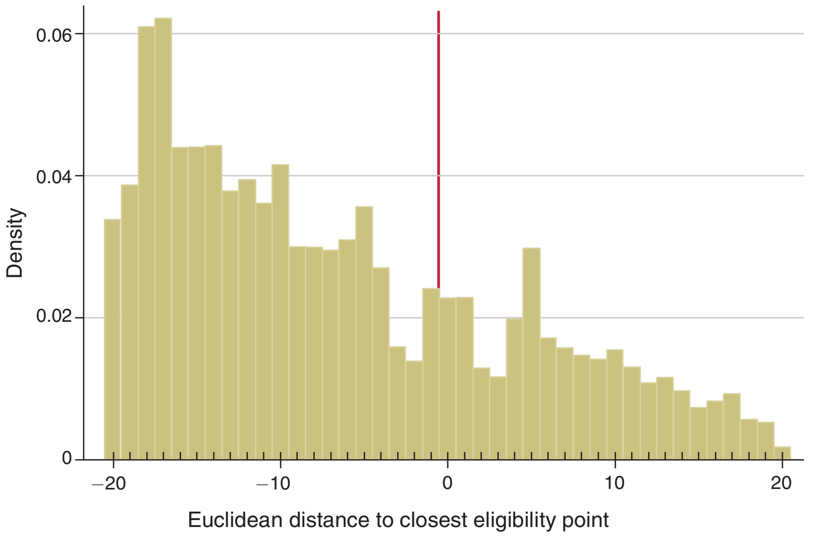


Figure 5. Distribution of Distances to Boundary

Table 2—Reduced-Form Estimates of Discontinuities in Preexisting (Fifth Grade) Student Characteristics

	Black (1)	Hispanic (2)	Female (3)	LEP (4)	Gifted in fifth grade (5)	Special education (6)
Above GT cutoff	−0.001	0.020	0.022	0.041	0.010	0.008
	(0.024)	(0.030)	(0.036)	(0.031)	(0.036)	(0.009)
Observations	4,102	4,102	4,102	4,101	4,102	4,102
	Free/ reduced-price lunch (7)	Stanford math (8)	Stanford reading (9)	Stanford language (10)	Stanford social studies (11)	Stanford science (12)
Above GT cutoff	−0.001	−0.149***	−0.017	−0.065	−0.042	−0.011
	(0.029)	(0.042)	(0.042)	(0.047)	(0.047)	(0.048)
Observations	4,102	4,075	4,075	4,071	4,074	4,073
	Any missing matrix data (13)	Teacher score (14)	Teacher points (15)	Enrolled (16)	Enrolled (free/ reduced-price lunch) (17)	Enrolled (non-free/ reduced-price lunch) (18)
Above GT cutoff	0.001	2.024	0.361	0.049**	0.043	0.063*
	(0.006)	(2.065)	(0.246)	(0.022)	(0.029)	(0.036)
Observations	4,102	4,097	4,097	5,368	3,425	1,943

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Regressions include a linear smoother with a slope shift above the cutoff. The sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between −15 and 15. Standard errors are robust to heteroskedasticity and clustered by fifth grade school. The estimation sample-students observed in LUSD two years after evaluation (seventh grade)—is used is for columns 1–15, while all fifth graders are included in columns 16–18. Regressions using the full set of evaluated students provides similar results and is provided in the online Appendix.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

the only covariate that is significant we believe this to be a spurious result. Nonetheless, since achievement is highly correlated over time we correct for this by providing results both with and without controls that include the lagged (fifth grade) dependent variable.

In column 13 we test whether there is any difference in whether a component of the matrix is missing, and find no such evidence. The next two columns address whether teachers manipulate evaluations for students at the qualification threshold.[32] If this were the case we would expect to find a discontinuity in the teacher scores, or in the matrix points the student gets from the teacher. We find no statistically significant discontinuity in either measure of teacher recommendation. Later we provide results for an additional teacher manipulation test that shows little impact on our estimates.

Finally, in columns 16 though 18 we test whether there is a discontinuous likelihood of being enrolled in the district two years after evaluation. Like Figlio and Page (2002) and Davis et al. (2010), we find that students identified as GT are about 4.9 percentage points (column 16) more likely to stay in the district, off a base retention rate of 76 percent in our estimation sample. Columns 17 and 18 suggest that this is more distinct for higher income students, though the two groups do not significantly differ. Nonetheless, there are two reasons that the impact of differential attrition on our estimates is likely to be negligible. First, as noted above, in Table 2 we only see differences across the discontinuity in math for those who remain in the district, and if anything the math estimate is in the opposite direction of what we would expect. If differential attrition were a problem we would expect to see significant differences in observables across the discontinuity. Further, in online Appendix Table 1 we provide estimates using the full 2007–2008 fifth grade cohort and find very similar results to those in Table 2. Second, we conduct bounding exercises to test the sensitivity of our estimates to attrition. The first analysis assigns each student who leaves LUSD after fifth grade the achievement score they received in fifth grade. We then estimate the reduced form RD model. The results are very similar to the baseline estimates we discuss below. In the second test we make the same imputation initially, but then progressively increase the (pseudo) seventh grade score for students below the cutoff who attrit until we reach statistical significance at the 5 percent level. This model indicates that the attritters below the cutoff would have to score, on average, between 0.15 and 0.26 standard deviations higher than their fifth grade scores to generate positive GT effects. This is a very large effect size, and thus, combined with the results in Table 2, it seems unlikely that differential attrition is affecting our estimates in an important way.

B. *Treatment on Marginal GT Students*

Because GT programs include a variety of services, we use our baseline RD model to illustrate how the educational environment for students changes when they are admitted to the GT program. Table 3 presents estimates from the regression-discontinuity GT model on several outcomes including peer achievement, where a student's peers are determined using fifth grade scores for other

[32] Although teacher recommendations are due before achievement scores are calculated, district officials informed us that many teachers submit their recommendations late.

Table 3—2SLS Estimates of Impacts of GT Services

	Peer math scores in math classes (1)	Peer reading scores in reading/English classes (2)	Peer language scores in reading/English classes (3)	Peer social studies scores in social studies classes (4)	Peer science scores in science classes (5)	Attends non-zoned GT magnet campus (6)	Attends zoned school (7)
Enrolled in GT	0.316***	0.274**	0.296**	0.266***	0.278**	0.154*	−0.043
	(0.117)	(0.118)	(0.111)	(0.098)	(0.108)	(0.085)	(0.081)
Observations	4,058	3,810	3,810	3,954	3,954	4,067	4,067

	Attends other non-zoned (8)	Number of core vanguard classes (9)	Enrolled in vanguard math (10)	Enrolled in vanguard English (11)	Enrolled in vanguard social science (12)	Enrolled in vanguard science (13)	Average teacher is GT certified (14)
Enrolled in GT	−0.111	1.250**	0.314**	0.290**	0.320**	0.320**	0.025
	(0.072)	(0.499)	(0.127)	(0.131)	(0.130)	(0.130)	(0.032)
Observations	4,067	4,090	4,058	3,813	3,954	3,954	3,905

	Average teacher has master's or higher degree (15)	Average teacher experience (16)	Average teacher has 0–2 years experience (17)	Average teacher has 3–10 years experience (18)	Average teacher has > 10 years experience (19)
Enrolled in GT	0.046	1.199*	−0.038	−0.031	0.069*
	(0.041)	(0.616)	(0.030)	(0.041)	(0.041)
Observations	3,905	3,905	3,905	3,905	3,905

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Controls for race, gender, economic disadvantage, LEP, prior gifted status, and lagged (fifth grade) dependent variable included. Also includes a linear smoother with a slope shift above the cutoff. Peers are defined by teacher—course ID—grade cells and peer achievement is measured using fifth grade test scores. Note there are no separate reading courses in seventh grade. The sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between −15 and 15. Standard errors are robust to heteroskedasticity and clustered by seventh grade school.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

seventh grade students in each grade-teacher-course cell,[33] school choices, enrollment in "Vanguard" (VG) classes with advanced curricula targeted to gifted students, and observable teacher characteristics. In columns 1 to 5 we find that peer achievement is measured to be between 0.27 and 0.32 standard deviations higher for GT students relative to non-GT students.[34] In column 6 we see that GT students are more likely to attend a school with a GT magnet program. Interestingly columns 7 and 8 suggest that most of this increase comes from students switching from other choice (magnet) schools rather than their zoned schools, though the estimates are

[33] Ideally one would like to use the actual classroom as the peer group. Unfortunately specific course section data are not available. To test the extent to which this is an issue, in online Appendix Table 2 we sort students within each cell into synthetic classrooms of at most 35 students under the assumption that students are tracked by their fifth grade achievement in the given subject (row 1) or randomly (row 2). With the exception of math in seventh grade the estimated changes in peer achievement are similar to those found in Table 5 under both assumptions.

[34] Reduced-form and first-stage results are available by request.

not significant. The table also shows that GT students are more likely to enroll in VG classes. If all GT students would have attended non-VG classes in the absence of GT qualification, the value in column 9 would be expected to equal four.[35] In fact the number, while significant, is much smaller at 1.25. The reason for this is that many non-GT students are permitted to take these VG courses, and in some cases GT students are permitted to take regular courses. Nonetheless, columns 10 through 13 show that on a per-course basis GT students are around 30 percent more likely to enroll in a VG class. Finally, in columns 14 to 19 we provide impacts on the characteristics of the students' teachers. Since some students have multiple teachers in each course, we calculate average teacher characteristics by first taking the mean within each course for each student and then averaging over all four of the tested subjects for each student. The estimates show that GT students are no more likely to have a GT certified teacher[36] or a teacher with an advanced degree, although their teachers have slightly more experience.

C. *Achievement Results Using Regression Discontinuity*

We present both graphical and econometric evidence that GT exposure for marginal GT students has no discernible effects on achievement test scores after a year and a half of GT program exposure. We subject our baseline analysis to a series of sensitivity permutations. Specifically, we examine variations in GT treatment intensity based on school specific variation. We also examine a spread of differences in alternatives for marginal students that are not declared GT. Finally, we explore the typical alternative specifications for the basic analysis. In no case are we able to find a comparison where the marginal GT students significantly out-perform their regular curriculum peers.

Figure 6 graphically presents the initial reduced-form results for each of the five achievement tests, showing little evidence of discontinuities at the eligibility boundary, with the exception of a dip for math. This is confirmed in panel A of Table 4, where there is a significantly negative effect in math. Given the negative estimate for lagged math in Table 2, we believe this is simply the result of the auto-regressive nature of achievement. Thus in panel B we include student level controls measured during fifth grade—prior achievement, race, gender, economic disadvantage, LEP status, and gifted status. Specifically, the third and fourth rows of each panel show the fuzzy RD estimates. This is our preferred model. We find in the first stage that being above our boundary is strongly associated with GT participation. The fourth row, however, shows all of the 2SLS estimates on achievement exam performance differences are close to zero. The 95 percent confidence intervals around the estimates rule out positive impacts of GT on marginal students of more than 0.11–0.18 standard deviations depending on the subject. The point estimates themselves, however, clearly suggest a zero effect for

[35] These include math, science, social studies, and English/language arts. We exclude reading as very few students in our sample take reading in seventh grade.

[36] While this may seem surprising, we note that in many schools the same teacher instructs both the more advanced non-Vanguard classes and the Vanguard classes.

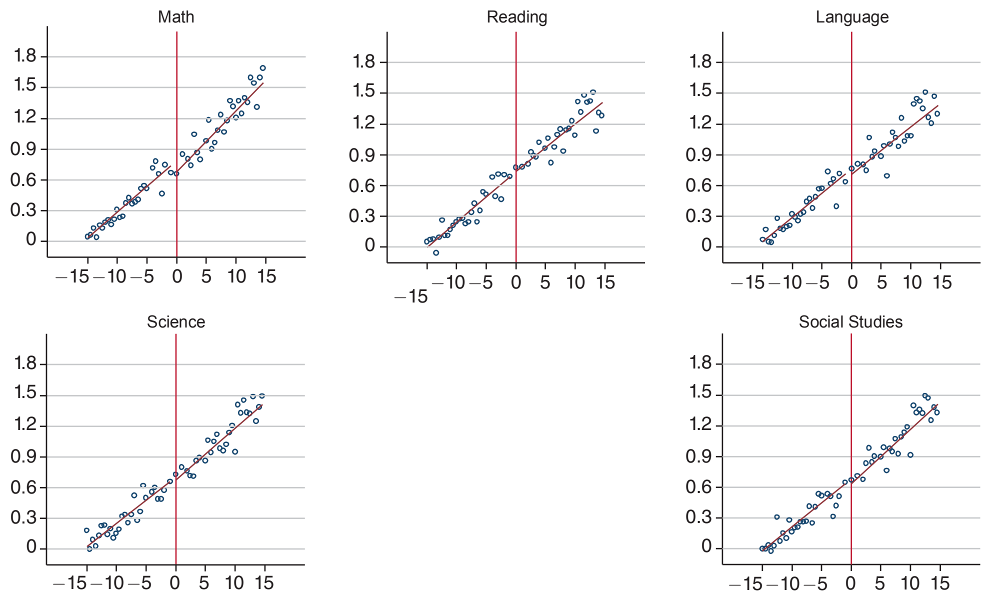


FIGURE 6. REDUCED-FORM EFFECTS ON ACHIEVEMENT IN SEVENTH GRADE BY DISTANCES TO BOUNDARY

Note: Achievement measured in standard deviation units within grade and year.

this rather broad, but nonetheless marginal group of GT students.[37] We further note that these estimates differ substantially from the OLS estimates on the same restricted sample shown in the first row of panel B. With the exception of reading, the OLS estimates are consistently larger than the 2SLS estimates and are significant for language, science and social studies.

Finally in panel C we conduct another test of manipulation of teacher scores. In this panel, for those students whose teacher score is pivotal (e.g. after accounting for all other matrix components, the teacher score determines eligibility), we replace the teacher points with imputed values based on scores in the other portions of the matrix.[38] This purges the Euclidean distance of variation from teacher points for any student where manipulation of the teacher component could affect qualification status. However, it adds error to the measure, thus reducing precision. Nonetheless, the estimates in panel C, while less precise, show results similar to those in panel B, further indicating that teacher manipulation of their portion of the matrix score is not a substantial concern.

[37] Online Appendix Table 3 shows that our results are robust to including only the lagged dependent variable as controls and to adding middle-school fixed effects. We also note that results are similar if we add elementary school fixed effects and are similar and more precise for sixth grade outcomes. These are available by request.

[38] Specifically, we estimate $Total\_Points_i = \alpha + \beta_1 Stanford\_Points_i + \beta_2 NNAT\_Points_i + \beta_3 Obstacle\_Points_i + \beta_4 Grade\_Points_i + \varepsilon_i$ for all students in the sample and capture the predicted values, $\widehat{Total\_Points}_i$. Thus we predict total points based on all components of the matrix excluding teacher recommendations. Then, for students where teacher points are pivotal, we calculate a synthetic value for "other points," which includes teacher points in our baseline estimates: $\widehat{Other\_Points}_i = \widehat{Total\_Points}_i - Stanford\_Points_i - NNAT\_Points_i$.

Table 4—Regression Discontinuity Estimates of Impact of Receiving G&T Services

Model	Independent variable	Stanford Achievement Test Math (1)	Reading (2)	Language (3)	Social studies (4)	Science (5)
Panel A. Baseline—no controls						
OLS (RD Sample)	Enrolled in GT	0.320***	0.391***	0.353***	0.379***	0.406***
		(0.085)	(0.084)	(0.068)	(0.078)	(0.083)
Reduced form	Above GT cutoff	−0.108**	0.000	−0.053	−0.036	−0.049
		(0.050)	(0.047)	(0.042)	(0.048)	(0.046)
2SLS—1st stage	Above GT cutoff	0.495***	0.497***	0.496***	0.495***	0.495***
		(0.051)	(0.051)	(0.051)	(0.051)	(0.051)
2SLS—2nd stage	**Enrolled in GT**	**−0.219****	**0.000**	**−0.107**	**−0.072**	**−0.100**
		(0.096)	**(0.095)**	**(0.082)**	**(0.093)**	**(0.091)**
Observations		4,053	4,055	4,054	4,053	4,050
Panel B. Baseline with individual controls						
OLS (RD Sample)	Enrolled in GT	0.064	0.045	0.078**	0.089**	0.097***
		(0.043)	(0.037)	(0.035)	(0.034)	(0.035)
Reduced form	Above GT cutoff	−0.019	0.024	−0.008	0.001	−0.013
		(0.038)	(0.034)	(0.033)	(0.042)	(0.041)
2SLS—1st stage	Above GT cutoff	0.507***	0.501***	0.501***	0.501***	0.500***
		(0.054)	(0.054)	(0.054)	(0.054)	(0.054)
2SLS—2nd stage	**Enrolled in GT**	**−0.037**	**0.049**	**−0.015**	**0.003**	**−0.025**
		(0.074)	**(0.068)**	**(0.066)**	**(0.084)**	**(0.084)**
Observations		4,023	4,025	4,020	4,022	4,018
Panel C. Using synthetic matrix scores (with controls)						
Reduced form	Above GT cutoff	−0.024	0.012	−0.030	0.004	−0.048
		(0.032)	(0.026)	(0.039)	(0.039)	(0.057)
2SLS—1st stage	Above GT cutoff	0.343***	0.344***	0.342***	0.341***	0.341***
		(0.040)	(0.040)	(0.041)	(0.040)	(0.041)
2SLS—2nd stage	**Enrolled in GT**	**−0.061**	**0.029**	**−0.093**	**0.017**	**−0.139**
		(0.094)	**(0.076)**	**(0.111)**	**(0.117)**	**(0.174)**
Observations		3,926	3,925	3,922	3,921	3,918

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Synthetic matrix scores replace matrix scores for students where a teacher recommendation could be pivotal (e.g., total points w/o the recommendation is fewer than ten away from the relevant cutoff) with the predicted value from a regression of total points on all components excluding the teacher points. See text for details. Controls for race, gender, economic disadvantage, LEP, prior gifted status, and lagged (fifth grade) dependent variable included in panel B. All panels include a linear smoother with a slope shift above the cutoff. Sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between −15 and 15. Standard errors are robust to heteroskedasticity and clustered by seventh grade school.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

In Tables 5A and 5B, we investigate whether there are heterogeneous impacts by certain characteristics of the middle schools the students attend. We caution that these results are merely suggestive as GT eligibility influences the school a student chooses to attend; hence these estimates could be biased.[39] Nonetheless, these

[39] Potential biases might arise due to endogenous residential location decisions or endogenous decisions to enroll in a nonzoned, e.g., magnet, school.

Table 5A—Regression Discontinuity Estimates of Impact of Receiving G&T Services Samples Split by Type of School Student Attends in Seventh Grade

Model	Independent variable	Subject: Math (1)	Math (2)	Reading (3)	Reading (4)	Language (5)	Language (6)
Panel A. Whether students attend their zoned school							
		Zoned	Not zoned	Zoned	Not zoned	Zoned	Not zoned
2SLS—1st stage	Above GT cutoff	0.510***	0.509***	0.503***	0.504***	0.505***	0.504***
		(0.067)	(0.066)	(0.067)	(0.065)	(0.066)	(0.067)
2SLS—2nd stage achievement	Enrolled in GT	−0.115	−0.016	−0.068	0.100	−0.030	−0.016
		(0.108)	(0.106)	(0.082)	(0.094)	(0.098)	(0.086)
Observations		1,885	2,104	1,885	2,106	1,886	2,100
Panel B. Whether schools have a GT magnet program							
		Has magnet	Has no magnet	Has magnet	Has no magnet	Has magnet	Has no magnet
2SLS—1st stage	Above GT cutoff	0.402***	0.518***	0.404***	0.512***	0.404***	0.512***
		(0.071)	(0.065)	(0.075)	(0.065)	(0.074)	(0.066)
2SLS—2nd stage achievement	Enrolled in GT	0.110	−0.071	0.003	0.052	0.144	−0.073
		(0.148)	(0.093)	(0.233)	(0.075)	(0.176)	(0.075)
Observations		905	3,118	907	3,118	905	3,115
Panel C. Whether school offers at least one VG class in subject and grade							
		Does not offer VG	Offers VG	Does not offer VG	Offers VG	Does not offer VG	Offers VG
2SLS—1st stage impact on GT status	Above GT cutoff	0.437***	0.532***	0.437***	0.527***	0.437***	0.528***
		(0.098)	(0.064)	(0.096)	(0.065)	(0.096)	(0.066)
2SLS—1st stage impact on VG class enrollment	Above GT cutoff	—	0.227***	—	0.211**	—	0.217***
		—	(0.078)	—	(0.079)	—	(0.076)
2SLS—2nd stage achievement	Enrolled in GT	−0.089	−0.027	0.128	0.014	−0.178	0.021
		(0.202)	(0.073)	(0.171)	(0.075)	(0.135)	(0.069)
Observations		1,111	2,904	1,110	2,907	1,110	2,902

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Controls for race, gender, economic disadvantage, LEP, prior gifted status, and lagged (fifth grade) dependent variable are included. All panels include a linear smoother with a slope shift above the cutoff. Sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between −15 and 15. Standard errors are robust to heteroskedasticity and clustered by seventh grade school.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

models are useful for providing insight into whether our baseline results are a general impact of GT or are due to heterogeneous effects that vary by the treatments available to the student. In the first panel we estimate differences by whether or not the student attends the school to which they are zoned. Panel B shows the different impacts by whether or not the student attends a GT magnet school. Finally, panel C shows differences by whether or not the school offers a Vanguard class in seventh grade in the given subject. The table shows that there is little evidence of program effects regardless of how we cut the data as only one estimate (out of 30) is significant at the 10 percent level, and it is negative.

Tables 6A and 6B continue the search for GT program effects by segmenting schools according to the intensity of underlying treatments. We examine three aspects of potential treatment intensity; the achievement scores of peers, the share of students that take advanced (Vanguard) classes, and the share of students in a

Table 5B—Regression Discontinuity Estimates of Impact of Receiving G&T Services Samples Split by Type of School Student Attends in Seventh Grade

Model	Independent variable	Social studies (1)	Social studies (2)	Science (3)	Science (4)
Panel A. Whether students attend their zoned school					
		Zoned	Not zoned	Zoned	Not zoned
2SLS—1st stage	Above GT cutoff	0.505***	0.501***	0.503***	0.502***
		(0.067)	(0.066)	(0.067)	(0.068)
2SLS—2nd stage achievement	Enrolled in GT	0.054	−0.055	−0.177	0.093
		(0.126)	(0.112)	(0.114)	(0.099)
Observations		1,887	2,101	1,885	2,099
Panel B. Whether schools have a GT magnet program					
		Has magnet	Has no magnet	Has magnet	Has no magnet
2SLS—1st stage	Above GT cutoff	0.394***	0.513***	0.399***	0.511***
		(0.075)	(0.065)	(0.071)	(0.065)
2SLS—2nd stage achievement	Enrolled in GT	−0.269	0.058	−0.189*	0.021
		(0.188)	(0.108)	(0.077)	(0.104)
Observations		905	3,117	905	3,113
Panel C. Whether school offers at least one VG class in subject and grade					
		Does not offer VG	Offers VG	Does not offer VG	Offers VG
2SLS—1st stage impact on GT status	Above GT cutoff	0.437***	0.527***	0.435***	0.526***
		(0.097)	(0.065)	(0.097)	(0.066)
2SLS—1st stage impact on VG class enrollment	Above GT cutoff	—	0.221***	—	0.234***
		—	(0.077)	—	(0.074)
2SLS—2nd stage achievement	Enrolled in GT	0.135	−0.020	−0.127	0.004
		(0.234)	(0.087)	(0.211)	(0.085)
Observations		1,111	2,903	1,109	2,901

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Controls for race, gender, economic disadvantage, LEP, prior gifted status, and lagged (fifth grade) dependent variable are included. All panels include a linear smoother with a slope shift above the cutoff. Sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between −15 and 15. Standard errors are robust to heteroskedasticity and clustered by seventh grade school.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

class that are identified as GT. To do this we first estimate the following model for each school individually:

$$(5) \quad Treatment_{ijt+k} = \delta_j + \gamma_j Above_{it} + \rho_{1j} Dist_{it} + \rho_{2j} Dist_{it} \times Above_{it} + \mu_{ijt+k},$$

where "Treatment" is the particular treatment described in each panel of Table 6 for student i in school j and year $t + k$. Then, for those middle schools with at least 30 students for whom "Dist" is between -15 and 15, we split the sample based on the estimated value of γ_j. The cutoffs were chosen to ensure that those schools above the cutoff averaged substantially greater treatment on average than those below. Thus for mean peer achievement, we split the sample by whether the estimated impact is greater or less than 0.2 SD, for Vanguard enrollment we use 20 percentage points, and for percent gifted in the teacher-grade-year cell we use 10 percentage

Table 6A—Regression Discontinuity Estimates of Impact of Receiving G&T Services Samples Split by Estimated Impact of Cutoff on Mechanism for Student's Seventh Grade School

			Subject					
			Math		Reading		Language	
Model	Dependent variable	Independent variable	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. School's RD impact on peer achievement in teacher-course-grade-year cell								
			> 0.2 SD	<= 0.2 SD	> 0.2 SD	<= 0.2 SD	> 0.2 SD	<= 0.2 SD
2SLS—1st stage	Enrolled in GT	Above GT cutoff	0.556*** (0.095)	0.498*** (0.072)	0.578*** (0.073)	0.478*** (0.068)	0.619*** (0.097)	0.491*** (0.067)
2SLS—2nd stage mechanism	Mean peer achievement	Enrolled in GT	0.666*** (0.122)	0.144 (0.101)	0.656*** (0.151)	0.171 (0.150)	0.615** (0.161)	0.258* (0.134)
2SLS—2nd stage achievement	Own achievement	Enrolled in GT	−0.068 (0.093)	−0.091 (0.102)	0.111 (0.106)	0.003 (0.085)	−0.019 (0.060)	0.010 (0.084)
Observations			1,390	2,364	739	2,759	603	2,891
Panel B. School's RD impact on vanguard class enrollment								
			> 0.2	<= 0.2	> 0.2	<= 0.2	> 0.2	<= 0.2
2SLS—1st stage	Enrolled in GT	Above GT cutoff	0.629*** (0.048)	0.447*** (0.083)	0.639*** (0.046)	0.452*** (0.079)	0.634*** (0.050)	0.454*** (0.080)
2SLS—2nd stage mechanism	Vanguard class enrollment	Enrolled in GT	0.768*** (0.090)	−0.048 (0.111)	0.771*** (0.098)	−0.033 (0.108)	0.768*** (0.099)	−0.015 (0.107)
2SLS—2nd stage achievement	Own achievement	Enrolled in GT	−0.031 (0.084)	−0.151 (0.108)	0.120 (0.074)	−0.039 (0.095)	0.017 (0.075)	0.031 (0.109)
Observations			1,419	2,335	1,170	2,331	1,171	2,326
Panel C. School's RD impact on gifted share in teacher-course-grade-year cell								
			> 0.1	<= 0.1	> 0.1	<= 0.1	> 0.1	<= 0.1
2SLS—1st stage	Enrolled in GT	Above GT cutoff	0.620*** (0.061)	0.475*** (0.074)	0.626*** (0.057)	0.469*** (0.075)	0.624*** (0.059)	0.470*** (0.076)
2SLS—2nd stage mechanism	Gifted share of cell	Enrolled in GT	0.349*** (0.043)	0.110* (0.059)	0.360*** (0.054)	0.093 (0.064)	0.356*** (0.057)	0.099 (0.064)
2SLS—2nd stage achievement	Own achievement	Enrolled in GT	0.012 (0.089)	−0.130 (0.092)	0.125 (0.084)	−0.017 (0.091)	0.024 (0.062)	0.013 (0.100)
Observations			1,005	2,749	982	2,516	983	2,511

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Controls for race, gender, economic disadvantage, LEP, prior gifted status, and lagged (fifth grade) dependent variable are included. The sample is split based on schools with 30 or more students, based on the estimated treatment in each school. All panels include a linear smoother with a slope shift above the cutoff. Sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between −15 and 15. Standard errors are robust to heteroskedasticity and clustered by seventh grade school.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

points. In each panel, the second row provides the average difference in treatment. Thus in panel A we see that GT students in schools in the odd columns experience increases in average peer quality of between 0.40 and 0.67 standard deviations while those in schools in the even columns see increases of only 0.14 to 0.28 SD. Thus, if peer achievement is an important input into these students' education production we would expect to see that the impact in odd columns is larger than those in even columns. Similarly for Vanguard enrollment, the odd columns show increases of close to 80 percentage points while there is no change for those in the even columns. Finally, for percent gifted in the cell in panel C, odd column schools see increases on the order of 30 percentage points while even column schools see increases of around 10 percentage points. In all three cases, we see no significant differences between

Table 6B—Regression Discontinuity Estimates of Impact of Receiving G&T Services Samples Split by Estimated Impact of Cutoff on Mechanism for Student's Seventh Grade School

Model	Dependent variable	Independent variable	Subject: Science (1)	Science (2)	Social studies (3)	Social studies (4)
Panel A. School's RD impact on peer achievement in teacher-course-grade-year cell						
			> 0.2 SD	<= 0.2 SD	> 0.2 SD	<= 0.2 SD
2SLS—1st stage	Enrolled in GT	Above GT cutoff	0.606*** (0.105)	0.492*** (0.068)	0.602*** (0.108)	0.491*** (0.069)
2SLS—2nd stage mechanism	Mean peer achievement	Enrolled in GT	0.497** (0.109)	0.262** (0.118)	0.400** (0.111)	0.278** (0.119)
2SLS—2nd stage achievement	Own achievement	Enrolled in GT	−0.080 (0.137)	0.017 (0.104)	−0.026 (0.099)	−0.053 (0.107)
Observations			619	2,999	617	2,998
Panel B. School's RD Impact on Vanguard Class Enrollment						
			> 0.2	<= 0.2	> 0.2	<= 0.2
2SLS—1st stage	Enrolled in GT	Above GT cutoff	0.619*** (0.053)	0.455*** (0.084)	0.616*** (0.054)	0.455*** (0.084)
2SLS—2nd stage mechanism	Vanguard class enrollment	Enrolled in GT	0.782*** (0.098)	−0.014 (0.113)	0.780*** (0.098)	−0.016 (0.114)
2SLS—2nd stage achievement	Own achievement	Enrolled in GT	0.066 (0.105)	−0.091 (0.106)	−0.046 (0.067)	−0.067 (0.145)
Observations			1,284	2,366	1,282	2,365
Panel C. School's RD Impact on Gifted Share in Teacher-Course-Grade-Year Cell						
			> 0.1	<= 0.1	> 0.1	<= 0.1
2SLS—1st stage	Enrolled in GT	Above GT cutoff	0.630*** (0.059)	0.467*** (0.076)	0.627*** (0.061)	0.467*** (0.077)
2SLS—2nd stage mechanism	Gifted share of cell	Enrolled in GT	0.304*** (0.037)	0.106* (0.059)	0.302*** (0.037)	0.105* (0.059)
2SLS—2nd stage achievement	Own achievement	Enrolled in GT	0.143 (0.105)	−0.112 (0.097)	−0.052 (0.079)	−0.056 (0.125)
Observations			974	2,676	972	2,675

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Controls for race, gender, economic disadvantage, LEP, prior gifted status, and lagged (fifth grade) dependent variable are included. The sample is split based on schools with 30 or more students, based on the estimated treatment in each school. All panels include a linear smoother with a slope shift above the cutoff. Sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between −15 and 15. Standard errors are robust to heteroskedasticity and clustered by seventh grade school.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

the more and less intensely treated students and, in fact, in no case are any estimates significantly different from zero.[40] Further, there is no clear pattern to the results with some estimates from higher intensity being negative and others positive, and similarly for low intensity schools. Thus, the results in Table 6 combined with those from Table 5 strongly suggest that our baseline findings in Table 4 reflect a general characteristic of the GT program, and are not due to heterogeneity in any observable treatment dimension.

[40] An alternative is to re-do these analyses with students' zoned school to avoid potential selection via endogenous school choice. These estimates are provided in online Appendix Table 4, and show very similar results.

The first nine rows of Table 7 report on sensitivity analyses using the preferred model from panel B of Table 4, and we find our estimates change little. We provide estimates with alternative functional forms for the smoothing variable, adding middle school fixed-effects, limiting the data to observations with no missing matrix components, exploring alternative cutoffs for GT, using different sized bandwidths, or conducting local linear regressions with optimal bandwidths determined by leave-one-out cross validation.[41] In only two cases, language under a cubic spline smoother (row 2) and language while restricting the bandwidth to -7 to 7 (row 5) are the estimates statistically significant at the 10 percent level and in no cases are they significant at the 5 percent level.

The next two rows of Table 7 explore alternatives to using our distance index as a forcing variable for the GT designation. Specifically, we split the sample between students who score high on the Stanford and NNAT exams, and hence are eligible for the 56 point cutoff (row 10), and students who score lower on these tests and hence are eligible for the 62 point cutoff (row 11). In these cases we use the raw matrix score instead of the Euclidean distance, but the analysis nonetheless yields results similar to the baseline model showing marginal GT students do not significantly out-perform those in the regular curriculum. Finally, in row (12) we estimate whether GT impacts vary with prior achievement in the tested subject. In this model we continue to use the discontinuity as an instrument for enrolling in GT, and interact the discontinuity with prior achievement to generate an instrument for the interaction of GT with prior achievement. We see no evidence of differential impacts by prior achievement. Online Appendix Table 5 provides 2SLS estimates of GT impacts in seventh grade for various student sub-populations, but we find little evidence of differences by gender, race/ethnicity, economic status or prior gifted status.[42]

IV. Estimating the Impact of GT Magnet Schools Using Lotteries

A. *Tests of the Validity of the Random Lottery Design*

Table 8 presents balancing tests for the lottery sample, including both all students who enter the lottery in fifth grade as well as the subset that remains in the district through seventh grade. The results indicate that the lotteries are conducted in a random way, as the ex ante baseline (fifth grade) sample has no significant coefficients on any of the twenty covariates we test. Further, using the ex post estimation (seventh grade) sample shows no significant differences between winners and losers except for math, which is significantly higher for winners at the 10 percent level. Although one significant result out of eighteen can be spurious, it is nonetheless possible that this is due to differential attrition between lottery winners and losers. Indeed, when we estimate the impact of winning a lottery on attrition by seventh

[41] We do not report that the results are similarly unchanged if we use elementary school fixed effects, but these results are available on request.

[42] The only exceptions are that students who are not eligible for free or reduced-price lunch exhibit a drop in math scores (significant at the 10 percent level) while a similar drop is also seen for student who were designated as "gifted" in elementary school (significant at the 5 percent level).

Table 7—2SLS Regression Discontinuity Estimates of Impact of Receiving G&T Services Specification Checks

	First stage	Stanford Achievement Test: Math	Reading	Language	Social studies	Science
	(1)	(2)	(3)	(4)	(5)	(6)
(1) Quadratic spline as smoother	0.427***	0.028	−0.016	0.126	0.151	0.079
	(0.063)	(0.128)	(0.126)	(0.130)	(0.146)	(0.137)
Observations	4,047	4,023	4,025	4,020	4,022	4,018
(2) Cubic spline as smoother	0.416***	0.183	−0.032	0.380*	0.286	0.097
	(0.071)	(0.269)	(0.188)	(0.200)	(0.200)	(0.222)
Observations	4,047	4,023	4,025	4,020	4,022	4,018
(3) Add middle school fixed effects	0.507***	−0.049	0.036	−0.004	0.001	−0.022
	(0.052)	(0.064)	(0.068)	(0.063)	(0.089)	(0.083)
Observations	4,047	4,023	4,025	4,020	4,022	4,018
(4) Limited to observations with no missing matrix data	0.507***	−0.031	0.059	−0.012	0.006	−0.022
	(0.053)	(0.076)	(0.069)	(0.063)	(0.087)	(0.078)
Observations	3,928	3,906	3,907	3,903	3,904	3,900
(5) Distance between −7 and 7	0.472***	0.097	0.010	0.194*	0.121	0.134
	(0.058)	(0.138)	(0.117)	(0.115)	(0.123)	(0.127)
Observations	1,800	1,789	1,794	1,789	1,789	1,786
(6) Distance between −11 and 11	0.474***	−0.033	0.030	0.044	0.018	0.039
	(0.060)	(0.088)	(0.085)	(0.088)	(0.111)	(0.089)
Observations	2,865	2,850	2,853	2,847	2,849	2,847
(7) Distance between −19 and 19	0.522***	−0.092	0.033	−0.017	−0.043	−0.035
	(0.050)	(0.059)	(0.057)	(0.064)	(0.072)	(0.063)
Observations	5,406	5,365	5,370	5,360	5,356	5,352
(8) Distance between −23 and 23	0.548***	−0.050	0.029	−0.030	−0.066	−0.042
	(0.048)	(0.051)	(0.048)	(0.053)	(0.058)	(0.043)
Observations	7,142	7,079	7,085	7,073	7,065	7,060
(9) Local linear regressions with rectangular kernel (reduced form)	—	0.140	−0.009	0.023	0.072	0.250
	—	(0.233)	(0.159)	(0.272)	(0.113)	(0.204)
Observations	—	1,075	1,078	708	2,043	1,074
Bandwidth for LLR (from leave-one-out cross validation)	—	5	5	3	8	5
(10) Limit to students who have 16 or more Stanford and 10 or more NNAT points	0.931***	0.031	0.009	0.093	−0.064	0.053
	(0.246)	(0.161)	(0.113)	(0.120)	(0.153)	(0.142)
Observations	1,639	1,630	1,632	1,628	1,629	1,629
(11) Limit to students who less than 16 Stanford or 10 NNAT points	0.861*	0.148	−0.089	0.165	0.064	−0.068
	(0.489)	(0.160)	(0.123)	(0.150)	(0.149)	(0.166)
Observations	2,408	2,411	2,411	2,410	2,411	2,407
(12) Interacting GT impacts with lagged achievement						
Enrolled in GT	0.525***	−0.032	0.073	−0.005	−0.001	−0.046
	(0.051)	(0.076)	(0.071)	(0.069)	(0.086)	(0.097)
Enrolled in GT × fifth grade achievement	0.722***	−0.006	−0.041	−0.016	0.008	0.035
	(0.047)	(0.032)	(0.034)	(0.028)	(0.041)	(0.049)
Observations	4,023	4,023	4,025	4,020	4,022	4,018

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Controls for race, gender, economic disadvantage, LEP, prior gifted status, and lagged (fifth grade) dependent variable included and a linear smoother with a slope shift above the cutoff except where noted. Sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between −15 and 15. Standard errors are robust to heteroskedasticity and clustered by seventh grade school.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 8—Balancing Tests for GT Magnet Lotteries—Covariates Measured in Fifth Grade

Sample	Asian (1)	Black (2)	Hispanic (3)	White (4)	Economic disadvantage (5)	Female (6)
Ex ante—baseline lottery	−0.030	0.030	0.041	−0.041	−0.035	−0.006
	(0.044)	(0.038)	(0.044)	(0.050)	(0.045)	(0.047)
Observations	542	542	542	542	542	542
Ex post— estimation sample	−0.027	0.041	0.042	−0.057	−0.050	−0.001
	(0.048)	(0.038)	(0.055)	(0.056)	(0.059)	(0.052)
Observations	437	437	437	437	437	437

Sample	At-risk (7)	Special education (8)	LEP (9)	Gifted (10)	GT magnet (11)	Total matrix points (12)
Ex ante—baseline lottery	−0.011	−0.019	−0.033	−0.028	0.035	0.243
	(0.010)	(0.017)	(0.022)	(0.035)	(0.030)	(0.926)
Observations	542	542	542	542	542	542
Ex post—estimation sample	−0.009	−0.015	−0.031	−0.024	0.055	0.909
	(0.011)	(0.023)	(0.027)	(0.047)	(0.045)	(1.173)
Observations	437	437	437	437	437	437

	Stanford Achievement Test				
Sample	Math (13)	Reading (14)	Language (15)	Social studies (16)	Science (17)
Ex ante—baseline lottery	0.027	0.073	−0.034	0.053	0.010
	(0.069)	(0.063)	(0.077)	(0.089)	(0.076)
Observations	540	541	539	540	539
Ex post—estimation sample	0.128*	0.100	−0.059	0.063	0.090
	(0.074)	(0.075)	(0.077)	(0.096)	(0.088)
Observations	437	437	436	437	436

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Lotteries for two schools were conducted in 2007–2008, hence, regressions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student won the lottery. Robust standard errors clustered by fifth grade school in parentheses.

grade we find that lottery winners are 11 percentage points less likely to attrite (standard error of 0.04).

We thus use the balancing results to inform our specification and analysis in three ways. First, as with the RD analysis, we present our results both with and without controls for lagged student scores and demographics. Second, to account for differential attrition, we re-weight the seventh grade estimation sample by the inverse of the predicted probabilities using a probit of attrition on fifth grade student characteristics.[43] Third, we estimate bounds on the impact of GT using a procedure

[43] Results of the probit regression are provided in online Appendix Table 6.

proposed by Engberg et al. (2010). The procedure uses observable characteristics to estimate the proportion of the sample that includes students of various types including those who are at risk of leaving LUSD if they lose the lottery. A generalized method of moments (GMM) estimator is used to generate upper and lower bounds. The upper bound assumes students at risk of leaving due to losing have achievement equal to the mean of students who stay and comply with the lottery results, while the lower bound assumes these same students score at the 95th percentile of the outcome distribution for all staying participants.[44]

B. *Results Using Randomized Lotteries*

Table 9 presents estimates showing the treatments received by GT students who win the lottery and attend the premier GT magnet schools. We find that lottery winners gain much stronger peers than losers. Using the attrition adjusted weighted results, peers in the magnet programs score between 0.7 and 1.2 SD higher than peers in the neighborhood GT programs. Further, the share of each student's teacher-grade-year cell classified as GT increases by 37 to 40 percentage points. On the other hand, there is no observable change to the curriculum, as the probability of taking advanced (Vanguard) classes in equal amongst winners and losers. Not surprisingly, we also see that lottery losers are much more likely than winners to remain in their zoned school. Finally, columns 15 to 20 show that lottery winners and losers have similar teacher characteristics using the results from the preferred weighted model.

Two-stage least squares estimates of the impact on student achievement from attending one of the two magnet GT programs are shown in Table 10.[45] Reduced-form estimates are provided in online Appendix Table 8.[46] We provide both inverse probability weighted estimates (rows 1 and 2) based on potential attrition, as well as unweighted estimates (rows 3 and 4). In rows 5 and 6 we provide upper and lower bounds to account for potential attrition bias using the Engberg et al. (2010) methodology.

This analysis shows the only case where measured achievement is found to improve due to GT treatment, as we find lottery winners attain a statistically significant 0.28 SD improvement in science based on the weighted analysis with controls. The unweighted sample with controls additionally shows a 0.14 SD improvement in language, but unlike science this result is not robust to the attrition correction.[47] Hence, we believe these estimates provide strong evidence of a lack of positive impact of attending a magnet on achievement other than in science.[48]

[44] That is, the upper bound assumes students at risk of leaving have average scores, while the lower bound assumes they are in the upper tail. These assumptions are those suggested by Engberg et al. (2010).

[45] In online Appendix Table 7, despite limitations due to small samples, we show there are very few discernible differences in impacts across student subpopulations.

[46] The first stage is always significant at the 1 percent level with point estimates of 0.47 (standard error of 0.11) for weighted and 0.57 (0.06) for unweighted regressions. Detailed first-stage results are available upon request.

[47] Results for sixth grade show somewhat larger, albeit still insignificant in the preferred model, impacts for math and language but not science. These are available by request.

[48] Ceiling effects is a potentially larger concern here than in the RD since the achievement levels of the lottery sample are higher. In online Appendix Figures 7–11 we provide distribution plots of raw scores on seventh grade exams by lottery winners and losers. Although the mass of achievement is further to the right than in the RD sample, the figures show there is nonetheless substantial room for improvement.

Table 9—Treatments in Seventh Grade for Students Attending a GT Magnet School Relative to Other GT Students

	Mean peer achievement (SD)					Vanguard class enrollment	
	Math in math class	Reading in English class	Language in English class	Social studies in soc class	Science in science class	Math	English
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2SLS—unweighted, controls	1.066*** (0.145)	0.659*** (0.149)	0.579*** (0.120)	0.794*** (0.123)	0.524*** (0.122)	0.056 (0.095)	0.065 (0.092)
Observations	440	436	436	439	439	440	440
2SLS—weighted, controls	1.164*** (0.179)	0.751*** (0.172)	0.686*** (0.143)	0.952*** (0.180)	0.659*** (0.166)	0.089 (0.100)	0.109 (0.095)
Observations	439	435	435	438	438	439	439

	Vanguard class enrollment		Share of teacher-course-grade-year cell who are GT				In zoned school in grade 7
	Social studies	Science	Math	English	Social studies	Science	
Model	(8)	(9)	(10)	(11)	(12)	(13)	(14)
2SLS—unweighted, controls	0.045 (0.090)	0.039 (0.091)	0.331*** (0.045)	0.356*** (0.050)	0.331*** (0.046)	0.331*** (0.046)	−0.348** (0.166)
Observations	440	440	440	436	439	439	438
2SLS—weighted, controls	0.096 (0.097)	0.083 (0.098)	0.378*** (0.062)	0.401*** (0.062)	0.373*** (0.063)	0.373*** (0.063)	−0.339** (0.137)
Observations	439	439	439	435	438	438	437

	Teacher characteristics (average across teachers and subjects)					
	GT certified	Master's or higher degree	Experience	0 to 2 years experience	3 to 10 years experience	> 10 years experience
Model	(15)	(16)	(17)	(18)	(19)	(20)
2SLS—unweighted, controls	0.001 (0.099)	0.143** (0.056)	2.217** (1.015)	−0.048 (0.071)	−0.061 (0.089)	0.108 (0.069)
Observations	439	439	439	439	439	439
2SLS—weighted, controls	−0.012 (0.089)	0.079 (0.066)	1.276 (1.272)	−0.060 (0.061)	0.025 (0.095)	0.036 (0.078)
Observations	438	438	438	438	438	438

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Lotteries for two schools were conducted in 2007–2008 hence all regressions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student is enrolled in a GT magnet program in seventh grade. Peers are defined by teacher—course ID—grade cells. Robust standard errors clustered by seventh grade school in parentheses. Results without clustering are similar and provided in the online Appendix. Weighted regressions are weighted by the inverse of the estimated probability of remaining in the data. See text for details. Controls include indicators during fifth grade for race, gender, special education, LEP, at-risk status, gifted, whether the student was enrolled in a GT magnet, and a lagged dependent variable.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

The bounding analysis in rows 5 and 6 of Table 10 confirm the results in row 2. Once again we see little to suggest that there is any substantial positive impact on math, reading, language, and social studies. For science, the upper bound of 0.34 is significant, although the lower bound is not different from zero.[49]

[49] We also show in online Appendix Table 9 that the results are similar, with the exception of science becoming smaller and insignificant, when we include nonlottery magnets in the comparison group (so that "GTMagnet" only

Table 10—Effect of Attending a GT Magnet School Relative to a GT Neighborhood Program

	Stanford Achievement Test				
Model	Math (1)	Reading (2)	Language (3)	Social studies (4)	Science (5)
(1) 2SLS—weighted, no controls	−0.266	−0.130	−0.060	−0.120	0.243
	(0.291)	(0.221)	(0.148)	(0.214)	(0.201)
Observations	436	437	435	436	436
(2) 2SLS—weighted, controls	−0.224	−0.018	0.001	−0.036	0.281**
	(0.171)	(0.172)	(0.114)	(0.136)	(0.130)
Observations	436	437	435	436	436
(3) 2SLS—unweighted, no controls	0.042	0.023	0.102	0.039	0.249**
	(0.178)	(0.103)	(0.065)	(0.083)	(0.114)
Observations	437	438	436	437	437
(4) 2SLS—unweighted, controls	−0.100	−0.058	0.142*	−0.032	0.208*
	(0.112)	(0.105)	(0.081)	(0.098)	(0.119)
Observations	437	438	435	437	436
(5) Engberg et al. (2010) bounds—upper bound	−0.019	−0.095	0.074	−0.064	0.344*
	(0.196)	(0.157)	(0.162)	(0.185)	(0.180)
Observations	437	438	436	437	437
(6) Engberg et al. (2010) bounds—lower bound	−0.353	−0.310	−0.207	−0.389	−0.013
	(0.251)	(0.192)	(0.215)	(0.249)	(0.248)
Observations	437	438	436	437	437

Notes: Achievement is measured in standard deviations of scale scores within grade and year. Lotteries for two schools were conducted in 2007–2008, hence all regressions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student is enrolled in a GT magnet program in seventh grade. Robust standard errors clustered by seventh grade school in parentheses. Results without clustering are similar and provided in the online appendix. Controls include indicators during fifth grade for race, gender, special education, LEP, at-risk status, gifted, whether the student was enrolled in a GT magnet, and a lagged dependent variable. Weighted regressions are weighted by the inverse of the estimated probability of remaining in the data. See text for details. In order to avoid slow convergence due to a very small portion of the sample being in special education or LEP, we drop those controls from the bounding analysis. Additionally, we do not cluster the standard errors on the bounding analysis due to inability of the estimator to converge.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

V. Discussion

Our analysis has examined two local average treatment effects of a GT program. We first use the eligibility cutoff to examine students that are marginally eligible for GT. We show that the marginally eligible students receive a different educational treatment than the marginally noneligible, as GT students take at least one more advanced class per year, and take classes with stronger peers. Nonetheless, we are not able to discern any positive achievement effects, even though the test we use should be able to detect such changes if they occur.

equals 1 if the student attends one of the two lottery schools instead of any of the eight GT magnets) and for alternative specifications of the definition of "Admitted" in the first stage.

The second treatment we examine is the intensity of GT programs, by comparing students that win to those who lose lotteries for admission to the two premier magnet GT middle schools. We demonstrate that students who enroll in the GT magnet programs take classes with substantially stronger peers and slightly more qualified teachers, although the number of advanced classes in which they enroll does not change. For these advanced students, we are able to find some evidence of improvement in science, although this result is not completely robust to our bounding estimates that account for differential attrition out of the school district.

In our view, there are at least three potential reasons we do not see achievement impacts on marginal students. First, it is possible that the GT program (or providing more intensive services) has a positive effect, but this effect is offset by additional support for students who do not qualify from parents. Such additional support could result from informal mechanisms, such as more time investment by the parent in the child's schooling, or more formal pathways such as additional tutoring or enrichment activities.

Second, the school district sets the boundaries for GT eligibility, and defines the GT curriculum (subject to state mandates). If the school district has a goal to keep as many students as possible, and if parents (and/or students) have a high demand for GT program participation, as is evidenced by Davis et al. (2010) and our own attrition estimates, it may be optimal for the district to expand participation as long as it does no educational harm. While understanding school district goals awaits further analysis, our RD results are consistent with this type of behavior. Further, the lottery results are consistent with a district that desires to maximize GT exposure intensity until there are no further educational gains. Clearly, further analysis is needed to make this conclusion, although we note that Abdulkadiroglu, Angrist, and Pathak (2011) and Dobbie and Fryer (2011) find similar results in their studies of elite high schools.

Third, given the strength with which peer effects have been found to operate in several different contexts, one would expect to find GT program achievement improvements from higher achieving peers even if there are no other effects (Angrist and Lang 2004; Hoxby and Weingarth 2006; Duflo, Dupas, and Kremer 2011; Lavy and Schlosser 2011; Imberman, Kugler, and Sacerdote 2012; Lavy, Paserman, and Schlosser 2012). Thus, in light of this research, our null results are somewhat surprising. One possible explanation is predicated on the fact that entering GT reduces a student's relative ranking within the class (Davis 1966). This could generate negative impacts through an invidious comparison model of peer effects where one's own performance falls with a reduction in one's position in the within-classroom achievement distribution (Hoxby and Weingarth 2006). We do not have sufficient data on student attitudes to investigate such a model, but we note that in an analysis of both the RD and lottery samples, not shown here but available by request, we find that course grades and class rank fall for treated students. An alternative explanation with similar empirical predictions is that teachers may target the material in their classes to the median or higher achieving students (Duflo, Dupas, and Kremer 2011), which could leave the students below the class average behind if the marginal students have difficulty absorbing the more advanced material. Even in the possible context of the preceding two paragraphs, therefore, it seems likely that some mechanism is causing impacts of peers in this portion of the achievement distribution to be ambiguous.

VI. Summary and Conclusion

In this paper, we identify the impact of providing gifted and talented services on student achievement. We exploit a unique universal evaluation for GT in fifth grade in a large urban school district. This allows for a regression discontinuity (RD) analysis of achievement gains for students on the margin of eligibility. We also use random lotteries for two elite magnet middle schools to assess achievement gains from more intense GT treatment. The combination allows us to view two separate portions of the ability distribution, as well as two types of services to GT students. The RD sample consists of students that are marginally eligible, although with a wider dispersion of abilities than usual because of the multidimensional eligibility criteria. The lottery sample includes students at the high end of the ability distribution who are higher achieving even than the average GT student.

To analyze these two groups, we first show that they both receive a different educational experience than the alternatives. In the RD sample, we find that GT students take at least one more advanced class each year, associate with stronger peers, and are placed with more GT students than marginally non-GT students. Further, GT teachers are slightly more experienced. For the GT magnet analysis we show the winners take classes with higher achieving peers, have a higher percentage of peers that are GT, and have slightly more educated and more experienced teachers.

Our analysis measures achievement gains via the Stanford Achievement Test, and we demonstrate that our samples are unlikely to be substantially affected by ceiling effects. Further, given Stanford test results are often used by school districts to determine GT eligibility, we believe it to be appropriate to evaluate these students using such an exam. Nonetheless, we acknowledge that GT services may affect outcomes other than test scores. Even so, test scores are the best data available to us for this particular evaluation.

We find that by seventh grade, marginal GT students show no significant change in achievement across all five subjects tested. Our analysis meets all of the standard validity tests for regression-discontinuity, and the results are robust to numerous specification checks. Further, we show that within the RD analysis, we see little evidence that impacts differ by the intensity of specific treatments, student characteristics, or the type of school attended. In the lottery analysis we also find that attending a GT magnet program has little impact on achievement. The exception is for science, which is positive and significant at 0.28 SD, although this result is somewhat sensitive to assumptions about the nature of differential attrition.

Our analysis is reduced form, so is unable to establish the path that leads to this null result. This presents a puzzle as we find large changes in peer quality amongst other changes generally considered to be positive. Thus, at the least the findings suggest that peer effects do not follow a standard monotonic model in all contexts. Further, there may be a limit in the ability of students to absorb additional material, which would put lower ability students at a disadvantage. The unanswered question is the extent to which these results are intentional, and are the outcomes of how the district maximizes its objectives relative to the constraints that it faces.

REFERENCES

Abdulkadiroglu, Atila, Joshua D. Angrist, and Parag A. Pathak. 2011. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." National Bureau of Economic Research (NBER) Working Paper 17264.

Angrist, Joshua D., and Kevin Lang. 2004. "Does Schooling Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review* 94 (5): 1613–34.

Argys, Laura, Daniel Rees, and Dominic Brewer. 1996. "Detracking America's schools: Equity at zero cost?" *Journal of Policy Analysis and Management* 15 (4): 623–45.

Barreca, Allan I., Jason M. Lindo, and Glen R. Waddell. 2011. "Heaping-Induced Bias in Regression-Discontinuity Designs." National Bureau of Economic Research (NBER) 17408.

Betts, Julian, and Jaimie Shkolnik. 2000. "The effects of ability grouping on student achievement and resource allocation in secondary schools." *Economics of Education Review* 19 (1): 1–15.

Bhatt, Rachana. 2011. "A Review of Gifted and Talented Education in the U.S." *Education Finance and Policy* 6 (4): 557–82.

Bhatt, Rachana. 2010. "The Impacts of Gifted and Talented Education." Unpublished.

Bui, Sa A., Steven G. Craig, and Scott A. Imberman. 2014. "Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Students: Dataset." http://dx.doi.org/10.1257/pol.6.3.30.

Card, David, and Alan B. Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy* 100 (1): 1–40.

Cicala, Steve, Roland G. Fryer, Jr., and Jörg L. Spenkuch. 2011. "A Roy Model of Social Interactions." National Bureau of Economic Research (NBER) 16880.

Clark, Damon. 2010. "Selective Schools and Academic Achievement." *B. E. Journal of Economic Analysis & Policy* 10 (1): Article 9.

Cullen, Julie Berry, Brian A. Jacob, and Steven Levitt. 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica* 74 (5): 1191–1230.

Davis, Billie, John Engberg, Dennis N. Epple, Holger Sieg, and Ron Zimmer. 2010. "Evaluating the Gifted Program of an Urban School District using a Modified Regression Discontinuity Design." National Bureau of Economic Research (NBER) Working Paper 16414.

Davis, James A. 1966. "The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men." *American Journal of Sociology* 72 (1): 17–31.

Dobbie, Will, and Roland G. Fryer, Jr. 2011. "Exam High Schools and Academic Achievement: Evidence from New York City." National Bureau of Economic Research (NBER) Working Paper 17286.

Duflo, Esther, Pascaline Dupas, and Micheal 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.

Engberg, John, Dennis Epple, Jason Imbrogno, Holger Sieg, and Ron Zimmer. 2010. "Evaluating Education Programs That Have Lotteried Admission and Selective Attrition." http://repository.cmu.edu/cgi/viewcontent.cgi?article=1443&context=tepper.

Epple, Dennis, Elizabeth Newton, and Ron Romano. 2002. "Ability tracking, school competition, and the distribution of educational benefits." *Journal of Public Economics* 83 (1): 1–48.

Figlio, David, and Marianne E. Page. 2002. "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Equality?" *Journal of Urban Economics* 51 (3): 497–514.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1): 201–09.

Hoekstra, Mark. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *Review of Economics and Statistics* 91 (4): 717–24.

Hoxby, Caroline, and Gretchen Weingarth. 2006. "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." Paper presented at the American Economic Association Annual Meetings, Boston, January 6–8.

Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote. 2012. "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review* 102 (5): 2048–82.

Jackson, C. Kirabo. 2010. "Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago." *Economic Journal* 120 (549): 1399–1429.

Lavy, Victor, M. Daniele Paserman, and Analia Schlosser. 2012. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in High and Low Achievers in the Classroom." *Economic Journal* 122 (559): 208–37.

Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics* 3 (2): 1–33.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–35.

Loveless, Tom, Steve Farkas, and Ann Duffett. 2008. *High-Achieving Students in the Era of NCLB*. Thomas B. Fordham Institute Report. Washington, DC, June.

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

Murphy, Patrick R. 2009. "Essays on Gifted Education's Impact on Student Achievement." PhD diss. Florida State University.

Naglieri, Jack A., and Margaret E. Ronning. 2000. "The Relationship between General Ability Using the Naglieri Nonverbal Ability Test (NNAT) and Stanford Achievement Test (SAT) Reading Achievement." *Journal of Pyschoeducational Assessment* 18 (3): 230–39.

Neal, Derek, and Diane Whitmore Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *Review of Economics and Statistics* 92 (2): 263–83.

Pop-Eleches, Cristian, and Miguel Urquiola. 2011. "Going to a Better School: Effects and Behavioral Responses." National Bureau of Economic Research (NBER) Working Paper 16886.

Reback, Randall. 2008. "Teaching to the rating: School accountability and the distribution of student achievement." *Journal of Public Economics* 92 (5–6): 1394–1415.

This article has been cited by:

1. David Card1 and Laura Giuliano2 1Department of Economics; 549 Evans Hall, #3880, University of California, Berkeley, Berkeley, CA 94720 and NBER (email: card@econ.berkeley.edu) 2Department of Economics, University of Miami, P.O. Box 248126, Miami, FL 33124 (email: l.giuliano@miami.edu) . 2016. Can Tracking Raise the Test Scores of High-Ability Minority Students?. *American Economic Review* **106**:10, 2783-2816. [Abstract] [View PDF article] [PDF with links]
2. Glenn Ellison, Ashley Swanson. 2016. Do Schools Matter for High Math Achievement? Evidence from the American Mathematics Competitions. *American Economic Review* **106**:6, 1244-1277. [Abstract] [View PDF article] [PDF with links]