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

Sa A. Bui[1], University of Houston

Steven G. Craig, University of Houston

Scott A. Imberman, University of Houston and NBER

Abstract

We identify the impact of gifted and talented services on student outcomes by exploiting a discontinuity in eligibility requirements and find no impact on standardized test scores of marginal students even though peers and classes improve substantially. We then use randomized lotteries to examine the impact of attending a GT magnet program relative to programs in other schools and find that, despite exposure to higher quality teachers and peers, only science achievement improves. We find that the relative ranking of students change, as do their grades, indicating that either invidious comparison peer effects or teaching targeting may be important.

October, 2011

[1] Department of Economics, 204 McElhinney Hall, Houston, TX 77204-5019. Correspondence should be made to Scott Imberman at simberman@uh.edu or Steven Craig at scraig@uh.edu. We would like to thank Aimee Chin, Dennis Epple, Jason Imbrogno, Chinhui Juhn, Brian Kovak, Jacob Vigdor and seminar participants at APPAM, AEFP, SOLE, Carnegie Mellon University, Georgetown University, University of Houston, and University of Maryland, 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.

I. Introduction

The focus of many school systems has become directed towards the low end of the achievement distribution as states have adopted accountability regimes and tried to meet the requirements of the federal No Child Left Behind Act (NCLB). As such, concerns have arisen that accountability pressures might have forced schools to shift resources away from high achieving students (Loveless, Farkas, and Duffett, 2008; Neal and Schanzenbach, 2010: Reback, 2008). Given these financial and regulatory constraints and the fact that over three million students in the US are classified as gifted and talented (GT), it is important to find out whether the services provided to these students are helpful. Hence, in this paper we provide what are, to our knowledge, the first credibly causal estimates of the effects of GT programs on high achieving students.

GT programs might help high achieving students through grouping with other high-achievers, and GT programs offer a variety of additional resources including specially trained teachers and a more advanced curriculum.[2] While early research finds that ability grouping is correlated with higher achievement, many of these studies are likely biased due to unobserved characteristics of students, such as motivation, that simultaneously lead students to be successful and to be grouped in high ability classrooms.[3] Recently, some research has tried to address the bias issue in ability grouping, but with mixed results (Argys, Rees and Brewer, 1996; Betts and Shkolnik, 2000; Duflo, Dupas and Kremer, 2011; Epple, Newlon and Romano, 2002; Figlio and Page, 2002;). Our analysis goes beyond these studies as it focuses on a group of students who are

[2] Conversations with district officials suggests that the GT curricula in the schools we studytend to include more detail and go more in-depth into topics, rather than cover the regular curriculum more quickly or add additional topics (increase breadth).
[3] See Kulik and Kulik (1997) for a review.

not well studied – high achievers – and the effectiveness of the services provided to them. As such, our work significantly expands understanding of GT programs, as we explicitly address program effectiveness using two unique strategies for overcoming bias - a regression discontinuity design embedded within the eligibility requirements, and an analysis of lotteries for entrance into two premier GT magnet programs.

Specifically, we utilize 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 non-verbal 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.

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 zone are randomly offered admission to the district's premier magnet schools. This allows us to examine achievement and attendance differences between students who win the lottery and attend the magnet GT schools, and those who lose the lottery and attend other "neighborhood" programs.[4] This analysis provides evidence on the impact of the extra inputs, including peers, because we provide evidence that the premier GT schools generate a more intensive treatment along observable measures than the neighborhood GT programs.

The combination of our two research strategies provides an unusually broad look at how GT programs affect student achievement. The RD approach uses students that are marginally eligible, although because there are several measurement dimensions for eligibility our work covers a broader array of students than would be typical in many RD applications. The lottery application captures the entire spectrum of student abilities, albeit with a relatively small sample. Each approach, however, helps to answer some of the questions from the other. The alternative program for students that are outside of the GT program (in the RD) is the "regular" track. The alternative for students who lose the lottery is to attend a "neighborhood" GT program in the local school.[5]

These features allow our work to inform several aspects of the economics of education literature. We demonstrate below that the treatment offered through the RD analysis provides curricular features not offered in the regular program and provides students with considerably stronger peers. On the other hand, the treatment in the lottery analysis has only small differences in curriculum from the alternative neighborhood programs. Students in the elite schools, however, are exposed to stronger peers and more effective teachers than those who attend neighborhood GT programs. Thus our analysis of the GT program assesses the impact of

[4] Students that lose the lottery 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.

[5] While students who lose the lottery also have the opportunity to attend a non-lottery magnet, the vast majority of those who stay in the district attend a neighborhood school program.

providing better peers and a more advanced curriculum in the RD analysis while the lottery analysis provides evidence on the impact of providing stronger peers and higher quality teachers.

To our knowledge, only Bhatt (2010) and Murphy (2009) specifically study the effect of GT programs on achievement, although Davis, Engberg, Epple, Sieg and Zimmer (2010) find that higher income parents are more likely to stay in public schools when their children are eligible for GT programs. Bhatt finds significant improvements in math achievement, although her instrumental variables methodology suffers from weak instruments. Murphy (2009) finds little math or reading improvement from being identified as GT, although these results may suffer from bias if trends in achievement determine program entry. Our work offers a wider scope of inquiry, and further offers the two distinct identification strategies. Thus, our study is the first to establish credibly causal estimates of the impacts of GT programs on student achievement.

Our RD analysis shows that students exposed to GT curriculum for the entirety of 6^{th} grade plus half of 7^{th} grade exhibit no significant improvement in achievement as measured by standardized test scores. This is despite substantial increases in average peer achievement on the order of one-fourth to one-third of a standard deviation, as well as being provided an advanced GT curriculum and the opportunity to attend a GT magnet program. We also find no significant improvements within student subgroups. While our estimates are local to the RD margin we note that, due to the multiple factors that go into the eligibility index, students near the discontinuity show wide variation in achievement levels, leading to a more generalizable interpretation than most RD designs. For example, students who score precisely at the eligibility threshold range from 45 to 97 national percentile rankings in reading and between 55 and 98 percentiles in math on the 5th grade exams.[6]

The lottery results that compare students in the two premier GT magnet schools to other GT students also show little improvement in 7th grade achievement with the exception of science scores. This is despite improvements in mean peer achievement on the order of one standard deviation, and higher quality teachers as measured by teacher fixed effects. We also note that, although the lottery analysis answers a different question than the RD, it has the advantage of evaluating the impact of receiving a more intense treatment for students who would be infra-marginal in the RD framework, as lottery participants tend to be from the upper part of the achievement distribution even within GT students.[7]

Peer effects are clearly a crucial dimension of the impact of GT programs given the large improvements in peers among GT students.. If peer effects follow a monotonic model whereby being surrounded by higher achieving students improves one's own achievement, as found in Imberman, Kugler and Sacerdote (forthcoming), better peers should lead to achievement gains. For a marginal GT student, however, the peer effect may not necessarily be positive. That is, a marginal GT student is likely to go from being near the top of the regular class to being near the bottom of the GT class. Further, even students in the middle of the GT distribution may experience a similar loss of relative rank in the magnet GT schools compared to neighborhood schools. In both contexts it is possible that an invidious comparison (IC) model applies, such as

[6] The inter-quartile ranges are 78 to 92 for reading and 83 to 94 for math. Additional measures for GT qualification include an intelligence test, grades, teacher recommendations and socioeconomic disadvantages.

[7] While the average GT student in 7th grade scores 1 standard deviation above the mean 7th grade student in the district, the lottery participants score 1.6 standard deviations above the mean.

proposed by Hoxby and Weingarth (2006), whereby students are demoralized by reductions in their relative ranking.

Further complicating the impact of peer effects in this context is that how the teacher targets her instruction (e.g. to the median, bottom or top student) can affect the benefits to the marginal student (Duflo, Dupas and Kremer, forthcoming). Further, Cicala and Fryer (2011) argue that the impact of moving students into an environment with higher achieving peers depends on where the student is in the achievement distribution. We therefore present suggestive evidence using both course grades and the within class ranking of students that indicates invidious comparison, or the effect of how the teacher targets classroom material, may be sufficiently important to balance out the other characteristics of GT programs that would be expected to increase achievement.

In a similar vein, if monotonic peer effects dominate these other potential mechanisms, selective schools should provide the largest positive impacts. Despite this common wisdom, however, Abdulkadiroglu, Angrist and Pathak (2011), Clark (2010), and Fryer and Dobie (2011) find that elite schools have little measured impact on achievement, although Jackson (2010) finds positive effects. Our lottery analysis of selective schools is distinguished from this work in part because it identifies the effect on an average student in the selective school as opposed to effects on marginal students as in these papers' RD analyses.

II. The Gifted and Talented Program in LUSD

LUSD is a large school district in the Southwestern US with over 200,000 students. The district is heavily minority and very low income, although the minority population is more heavily Hispanic. Panel A of Table 1 shows 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 cognitive and non-cognitive outcomes than non GT students.[8] In order 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 entry into GT in 2008-09 is provided in Figure 1. The matrix converts scores on standardized tests – Stanford Achievement Test for English speaking students and the Aprenda exam for a subset of Spanish speaking students with limited English proficiency – scores on the Naglieri Non-verbal Abilities Test (NNAT), average course grades, teacher recommendations, and indicators for socio-economic status into an index score we call "total matrix points."[9]

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.[10] Alternatively, students can qualify by having 62 total matrix points regardless of Stanford, Aprenda and NNAT scores. During 5th grade all students are evaluated for GT, including those who participated in the GT program in elementary school.[11] This

[8] Schools in LUSD also have a monetary incentive for attracting gifted students as LUSD provides a funding boost of 12% over the average daily allotment for a regular student.

[9] For socioeconomic status, students get 5 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 3 point bonus.

[10] Students can reach 16 points from the Stanford Achievement Tests through different combinations of scores on four subjects. For example a student who is in the 90th percentile in math and the 80th percentile in reading will qualify regardless of science and social studies scores. Alternatively a student could meet this requirement by scoring in the 80th percentile in all four exams. See Figure 1 for details on the score to points conversions. For the Naglieri test a score of 104 (no percentiles are given) would be equivalent to 10 matrix points.

[11] Elementary students must re-qualify in 5th 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 the GT program if they perform poorly. Those who do not qualify in 5th grade have the opportunity to be re-evaluated in later years at their school's discretion. All students are also evaluated for GT services in kindergarten, but unfortunately the matrix data was incomplete prohibiting us from evaluating the GT program in elementary schools.

selection framework allows us to model qualification across the eligibility boundary by using a fuzzy RD methodology. Specifically, while all students who meet the requirements above qualify, not all end up being classified as GT because parents are allowed to opt-out of the program, or students may enroll and then withdraw.[12] Further, some who do not initially meet the requirements later become identified as GT. This is mainly because students qualify for entry in 7th grade after not qualifying in 6th, parents appeal the recorded matrix scores by submitting an alternative standardized test taken within the prior 12 months, or data is added later.[13]

Table 1 also shows means from the lottery sample in panel B. The students in the lottery are significantly stronger academically than the average GT student in panel A. These differences allow an interesting contrast between the RD sample, which consists of marginal GT students, and the lottery sample, which consists of relatively strong GT students. For example, lottery participants in Panel B are shown to score about 1 standard deviation on standardized tests in 5th grade than the mean overall GT student from Panel A. [14] Lottery participants are also more likely to be white, and not on subsidized school lunch. A key element from the lottery sample is the attrition rate. Of the 542 students that enter the lottery, 18.8% are not in the school district by 7th grade, and in fact, most of these students leave the sample before 6th. Because the leavers are different than the lottery winners as shown in the last two columns, there is potential for attrition bias in the lottery sample. We address potential attrition bias in two ways – by

[12] 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. However, this is very rare as only 2 of the 41 traditional middle schools in LUSD had no GT students in 7th grade in 2009-10.

[13] Below we show that the missing matrix components do not appear to substantially influence our results. Additionally, results for 6th grade show a similar sized likelihood of being GT to the left of the boundary, suggesting that qualification in 7th grade rather than 6th is not an important factor.

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

reweighting the sample to look like the pre-lottery sample on observables, and through the use of a new bounding analysis proposed by Engberg, Epple, Imbrogno, Sieg and Zimmer (2010).[15]

III. Model and Specification

1. GT Program Evaluation Using Regression Discontinuity

The objective of the RD analysis is to estimate a local average treatment effect of providing gifted services to students who are on the margin of qualifying for these services. Figure 2 shows GT identification two years after evaluation (7th grade) as a function of the students' matrix points. The gradual increase up to 28% at the first cutoff (of students with a matrix score of 56) reflects missing matrix components, qualifying in 7th grade and the district's appeals 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% at the second cutoff (62 matrix points). After reaching the second cutoff, GT enrollment slightly increases further to 82 percent.

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 into three-dimensional space as shown 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 are eligible for GT while those below or behind it are ineligible. We then take the Euclidean distance from each student's total matrix points to the closest integer

[15] We also test for selective attrition in the RD sample but find little evidence to suggest it is a problem there.

combination on the surface.[16] The resulting value, which we call the distance to the qualification threshold, equals zero if the student just barely qualifies for GT. Figure 4 shows GT enrollment as a function of Euclidean distance from the qualification threshold. Students just below the cutoff have a 25% likelihood of being in GT, however students just above the threshold have a likelihood of approximately 79 percent.[17]

Since qualification for GT via the observed matrix score does not translate perfectly with enrollment in GT due to appeals, substitute exams, opt-outs, and missing data, our estimation strategy uses a "fuzzy RD" model where we conduct 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). For most of this paper we will use ten distance units below and above the cutoff for our bandwidth since the relationships between distance and the achievement outcomes are close to linear over this range, allowing us to use a linear smoother. We show later that our results are not sensitive to the choice of bandwidth. Hence, we estimate the following two-stage least squares (2SLS) model:

$$(1)\ GT_{i,t+k} = \delta + \gamma Above_{it} + \rho_1 Distance_{it} + \rho_2 Distance_{it} \times Above_{it} + \boldsymbol{\Omega X_{it}} + \mu_{ijt+k}$$

$$(2)\ Y_{i,t+k} = \alpha + \beta GT_{i,t+k} + \lambda_1 Distance_{it} + \lambda_2 Distance_{it} \times Above_{it} + \boldsymbol{\Phi X_{it}} + \varepsilon_{ijt+k}$$

where $Above_{it}$ is an indicator for whether student *i* in year *t* has a distance measure at or above the cutoff, *Distance* is the Euclidean distance of the student's matrix score to the eligibility cutoff, and *X* is a set of pre-existing (5th grade) observable characteristics which includes the 5th

[16] The Euclidean distance is measured as $Distance_i = \sqrt{(Stanford_i - Stanford_s)^2 + (NNAT_i - NNAT_s)^2 + (Other_i - Other_s)^2}$ where *i* refers to the student's own score and *s* refers to the closest integer combination on the surface. We thank Jake Vigdor for first suggesting this method to us. This process allows us to simultaneously consider all three sources of matrix points (see Fig 1), without materially altering our estimation results.

[17] Note that by construction the distance measure has an empty mass between 0 and 1 and -1 and 0 since the smallest distance to another integer point is 1.

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 test scores, attendance, or disciplinary infractions in that year. Since students are tested in January of each year, we focus on outcomes in the second year after evaluation (7th grade) as assessment in the first year will only provide five months of program exposure, although we provide estimates for 6th grade outcomes in the online appendix.

2. GT Magnet Evaluation Using School Lotteries

LUSD has two middle schools with GT magnet programs that are over-subscribed, and as a result the district uses lotteries to allocate available spaces.[18] While the losers of the lottery still 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 and focus on advanced curricula.[19]

Our analysis compares the performance of students who win the lottery and attend one of the magnet GT programs to those who lose the lottery and either attend a neighborhood GT program in the district, a magnet school based on a different specialty, or a charter school. Hence in the lottery sample we estimate the following 2SLS model conditional on applying for admission to a magnet program with a lottery:

$$(3) GTMagnet_{ijt+k} = \delta + \gamma Admitted_{ijt} + \mathbf{\Omega} \boldsymbol{X_{it}} + \nu_j + \mu_{ijt}$$

$$(4) Y_{ijt+k} = \alpha + \beta GTMagnet_{ijt+k} + \mathbf{\Phi} \boldsymbol{X_{it}} + \eta_j + \epsilon_{ijt}$$

[18] There are 8 middle schools with GT magnet programs in total (out of 41 traditional middle schools), but only two 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 six GT magnet programs, and the remainder attend a neighborhood GT program . On the other hand, of the 265 lottery winners, only 3 attend one of the other six GT magnets in 7th grade.

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

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. Finally, since each school holds separate lotteries we include ν_j and η_j in the model as lottery fixed-effects.[20]

One caveat to the lottery is that students with an older sibling in the school are exempted from the lottery and automatically given admission. Unfortunately, LUSD was unable to provide data on siblings, but we believe siblings have a negligible impact on our estimates. First, all siblings need to apply and qualify for GT based on their matrix points. Second, we will show our lottery sample is very well balanced 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.

IV. Regression Discontinuity Estimates of GT Impacts

1. Data

Our data consists of the administrative records from 2007-08 to 2009-10. While we have data for universal assessments conducted in 2006-07, 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 in 2007-08, the second year of the mandatory GT assessment, and examine outcomes through the 2009-10 school year. For outcomes we use scale scores standardized

[20] Since we focus only on one cohort, 5th graders in 2007-08 (who are in 7th grade in 2009-10), there is a single lottery fixed-effect indicator in each regression. Models with 6th grade outcomes, and hence two cohorts of students, have three indicators. Also note that coding *GTMagnet* for students who attend a non-lottery GT magnet as zero instead of one has no effect on the results.

across LUSD within grade and year on the Stanford Achievement Test, as well as attendance rates and counts of disciplinary infractions warranting an in-school suspension or more severe punishment. 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.[21] After restricting the sample to a 20 unit band around the cutoff, we look at achievement of approximately 2,600 students in one 7th grade cohort and 5,500 students in two 6th grade cohorts who were evaluated for GT in 5th grade.[22]

2. Tests of Validity of RD Design

A primary concern with any regression-discontinuity analysis is the potential for manipulation of the forcing variable (qualification for GT) that determines treatment. Such manipulation could bias the results if it is correlated with outcomes (Lee and Lemieux, 2010). In Figure 5, we provide density plots around the discontinuity showing differences in density around the discontinuity are similar in size to changes at other parts of the distribution, suggesting that manipulation is unlikely to be occurring.[23]

In Table 2 we provide tests of discontinuities in pre-existing (5th grade) student characteristics.[24] We find no discontinuities in columns 1-14 for race, gender, LEP status, prior

---

[21] While some LEP students are evaluated using the Aprenda exam, a Spanish-language alternative to the Stanford Achievement Tests, only 0.5% of 5th grade students in LUSD take only the Aprenda exam and hence have no Stanford scores. Thus, we drop students who only have Aprenda scores.

[22] Within this bandwidth the total matrix scores have an inter-quartile range of 48 to 65 with a minimum of 39 and a maximum of 79.

[23] 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 16) 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 manipulation. In both cases the test is statistically insignificant. We also provide graphical evidence on the distribution of matrix points in Online Appendix Figure 1.

[24] A related concern is that laid out by Barreca, Guldi, Lindo and Waddell (2010) that heaping in running variables could lead to biased estimates if the heaps are correlated with unobservables and the bandwidths are small enough so that heaps are concentrated only on one side of the cutoff. We do not find any evidence of heaping in the matrix scores. Nonetheless, by construction some heaping will occur in the transformation from matrix scores to Euclidean

gifted status, special education status, eligibility for free or reduced-price lunch, disciplinary infractions, attendance rates, and achievement with the exception of math.[25] Given that math is 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 (5th grade) dependent variable.

In column (15) 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 teacher evaluations. The concern is that if teachers know a student is short of the qualification threshold, evaluations may be manipulated to qualify certain types of students.[26] 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. Nonetheless, below we provide an additional specification test to check for potential bias from teacher manipulation through their recommendations.

Finally, in columns (18) though (20) we test whether there is a discontinuous likelihood of being enrolled two years after evaluation. Given that Davis, et al. (2010) find evidence that high income students are more likely to stay in public schools if identified as GT, we check if such a phenomenon occurs in LUSD. Additionally, a discontinuity in attrition would be a marker

---

distances. This is not a problem in our context, however, as our bandwidths are wide enough to include substantial observations both at heaping and non-heaping points on both sides of the cutoff. We will also show later that our results are quite robust to choice of bandwidth.

[25] Online Appendix Figures 2 – 4 provide graphical representations of these results. Tests that do not condition on appearing in the data in 7th grade are similar and are provided in Online Appendix Table 1, along with tests using the 6th grade sample. These are also similar for all measures except for females which shows a small but statistically significant increase. In Online Appendix Table 2 we provide estimates without clustering of standard errors. These show no change in the significance levels of the estimates.

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

for potential attrition bias in the estimates. Nonetheless, we find no statistically significant change in the likelihood of enrollment at the discontinuity regardless of economic status. Hence, given these results and those described above we see little evidence that GT qualifications were manipulated in a way that would violate the assumptions underlying the RD methodology.

3. Results

Figure 6 presents the initial reduced-form results for three of the five achievement tests, and Figure 7 for the other two. These achievement test results are from 7th grade, and thus reflect about a year and a half of GT program exposure. We use a bandwidth in our regressions that includes students within ten distance units of the GT qualification boundary.[27] The coefficient estimates presented in the first panel of Table 3 as well as the means provided in Figure 5 show that there is no improvement in reading or language Stanford scores after a year and a half of GT participation, and that there is a negative and significant point estimate for math without individual controls.[28] Figure 6 confirms the findings in columns (4) and (5) of Table 3 of no discernible impacts on achievement in social studies or science.

Panel B of Table 3 provides estimates from our preferred specification of equation (2) that contains student level controls measured during 5th grade, including the lagged dependent variable, race, gender, economic disadvantage, LEP status, and gifted status. In this panel, all of 2SLS estimates are close to zero while math, reading and social science are negative and all t-statistics are below one. Drawing 95% confidence intervals around the estimates, we can rule out modest positive impacts of GT on marginal students of more than 0.06 standard deviations (sd)

[27] Online Appendix Figure 5 shows that our Euclidean distance measure correlates very well with total matrix points.

[28] Note that 89% of students in GT in 7th grade are also in GT in 6th grade and, conditional on remaining in LUSD, 4% of students who are in GT in 6th grade are not in 7th grade.

in math, 0.09 sd in reading, 0.15 sd in language, 0.13 sd in social studies and 0.23 sd in science. The point estimates themselves, however, clearly suggest a zero effect.[29]

In columns (6) and (7) we examine impacts on non-cognitive outcomes, disciplinary infractions and attendance rates. While there is no effect on disciplinary infractions, we do find a marginally significant negative effect on attendance rates. The drop in attendance rates of 1.1 percentage points is equivalent to attending school two fewer days in a 180 day school year. As we demonstrate below, however, this estimate is sensitive to the specification.

Panel C presents results that correct for the possibility of teacher manipulation. Even though our earlier tests did not suggest a problem, we are especially concerned because an administrator acknowledged that time deadlines are lax for recommendations, leaving opportunities for teachers to "top up" the scores of marginal GT students. We address this possibility by replacing a student's matrix points with a synthetic matrix score if the points from a teacher recommendation are potentially pivotal. That is, for students whose other matrix components place them within 10 points of the cutoff, the teacher recommendation, with a maximum of 10 points, is potentially determinative.[30] Thus for these students we replace their total matrix score with the predicted value from a regression using the full 5th grade sample of total matrix points on all matrix components excluding teacher points:

---

[29] Online Appendix Table 3 provides results for 6th grade. These are similar to those for 7th grade and are more precise due to the addition of an extra year of data. For social studies we can rule out effects of 0.15 sd while for other tests we can rule out impacts of 0.09 sd and higher. Appendix Table 4 provides results with the lagged dependent variable but without the other covariates. These are similar to the results in Panel B of Table 3. Finally Appendix Table 5 shows the results to be robust to the inclusion of middle school fixed-effects.

[30] While teacher recommendations would not be pivotal for students within 10 points above the cutoff, only adjusting scores on one side of the cutoff could introduce bias, particularly if teachers only have information on some of the components at the time they make their recommendations. Hence, we replace scores within 10 points above the cutoff with the synthetic scores as well.

$$(5) TotalPoints_i = \alpha + \beta StanfordPoints_i + \gamma NNATPoints_i + \delta ObstaclePoints_i + \eta GradePoints_i + \epsilon_i.$$

where *TotalPoints* is the student's final score on the GT qualification matrix, *StanfordPoints* are the number of matrix points received from performance on Stanford Achievement Tests, *NNATPoints* are matrix points from the non-verbal abilities test, *ObstaclePoints* are matrix points from socioeconomic status, and *GradePoints* are matrix points from the student's average grades in 5th grade. We convert these "synthetic matrix points" to Euclidean distances from the eligibility surface, thus purging the teacher component and any potential manipulation from the matrix scores.

The results using the synthetic scores including all controls used in panel B are provided in panel C. Since we are essentially adding measurement error to the first stage, the cutoff instruments are considerably weaker. Nonetheless they remain highly significant indicating that a discontinuity remains.[31] The 2SLS results show negative and insignificant effects for math, reading, language and social science while science estimates, although positive, are very close to zero. Discipline results are also similar to those in panel B. On the other hand, attendance impacts turn positive, albeit statistically insignificant. Given the potential influence of attendance on teacher recommendations, further RD analyses on attendance should be interpreted with caution. Nonetheless, since the results for all other outcomes are consistent with the estimates in panel B, the baseline model with controls is our preferred specification.[32]

[31] In Online Appendix Figure 6 we provide a graph of the first stage for the synthetic teacher scores. The point estimate for the first stage is 0.23 with a standard error of 0.04.

[32] We also estimated models that drop all students where their entry into GT could be impacted by teacher scores. While the estimates were very imprecise results were qualitatively similar to those in panel C of Table 3.

To test for heterogeneity in program impacts across student characteristics, Online Appendix Table 6 provides 2SLS estimates for 7th grade for various student sub-populations. In general, we find little evidence of differences by gender, race/ethnicity, economic status or prior gifted status. The only distinction is that the attendance estimates are more negative for women and black students and positive for white students, but we are cautious in drawing interpretations from this given the results in panel C of Table 3.

In Table 4 we test the sensitivity of our RD estimates in Table 3 to model specification using our preferred model with controls. We find that our estimates hold regardless of whether we add middle school fixed-effects, limit the data to observations with no missing matrix components, use smaller or wider bandwidths, or conduct local linear regressions with optimal bandwidths determined by leave-one-out cross validation. We also estimate models where, instead of using the distance index, we restrict the samples to students who score high on Stanford and NNAT and hence are eligible for the 56 point cutoff (row 5) and students who score low on these tests and hence are eligible for the 62 point cutoff. In these models instead of Euclidean distance, we use the raw matrix score as the forcing variable in our regressions. In both subsamples, the results are similar to our baseline model.

When we use a quadratic smoother as the functional form, however, the estimates show significant improvements in language and science achievement scores. These results become insignificant using a cubic smoother due to larger standard errors. Further inspection, however, suggests that this estimate is being driven by excessive curvature at the discontinuity. Online Appendix Figures 7 and 8 show that the quadratic estimates are mainly driven by random variation at the discontinuity and hence tend to overestimate what appears to be the true impact. As such, we believe a linear smoother captures the correct estimates.

A further concern is that the lack of positive effects may be due to top-coding of the exams. Since GT students are high-achievers many of them may not be able to exhibit growth on achievement tests as they are very close to answering every question correctly. To address this, in Online Appendix Figures 9 – 13 we provide distribution plots of raw scores on each of the 7th grade Stanford Achievement Tests for students with Euclidean distances between -10 and 10. Not surprisingly for these marginal GT students, in all cases the mass of the distribution is centered far from the maximum score. For example in math the modal score is 62 out of 80 while it is 67 out of 84 for reading, leaving substantial room for improvement.

Another potential reason for not finding an effect of GT services on student outcomes is that there may be little treatment on students. In Table 5, however, we illustrate the extent to which entering GT generates a measureable treatment. We estimate the impact of GT on peer achievement, where a student's peers are determined by other students in a grade-teacher-course cell,[33] school choices, teacher quality and enrollment in "Vanguard" classes which are pre-Advanced Placement classes with advanced curricula targeted to gifted students.

Teacher quality is measured through the use of teacher fixed-effects using a procedure in the spirit of Kane and Staiger (2008). Specifically, we use teacher-student linked achievement data for LUSD middle school students from 2007-08 to 2009-10 to estimate the following model:

$$(6)\ Y_{ijkt} = \alpha + \delta Y_{ijkt-1} + \boldsymbol{\Omega X_{ijkt}} + \boldsymbol{\Phi Z_{ijkt}} + \gamma_{\mathrm{k}} + \delta_{\mathrm{j}} + \epsilon_{\mathrm{ijkt}}$$

[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 7 we sort students into synthetic classrooms of at most 35 students under the assumption that students are tracked by their 5th grade achievement in the given subject (row (i)) or randomly (row (ii)). With the exception of math in 7th grade the estimated change in peer achievement is similar to those found in Table 5 under both assumptions.

where Y is achievement in a given subject, and X is a set of student level controls for economic status, gender, race/ethnicity, special education status, LEP status, and grade by year fixed effects. Z is a set of controls for mean peer achievement (defined at the grade-course-teacher level) in each Stanford Achievement Test. Finally, δ_j is a set of school fixed effects and γ_k is a set of teacher fixed effects.[34] Kain and Staiger (2008) show that a similar framework closely replicates the results from a randomized allocation of students to teachers.[35]

Table 5 shows in columns (1) to (5) that peer achievement is between 0.24 and 0.35 standard deviations higher for GT students relative to non-GT students.[36] The table also shows that GT students are more likely to enroll in Vanguard classes and attend a GT magnet program. Interestingly, most of the school switching does not appear to come from students leaving their zoned school for GT magnets. Rather, students appear to move from schools other than their zoned school – mostly non-GT magnets – to the GT magnets. Finally, and perhaps surprisingly, GT students do not appear to get assigned better teachers as measured by teacher fixed effects. This may be due to the fact that in many schools both GT and non-GT students can access advanced classes taught by the same teacher. Nonetheless, the change in peers and the increase in enrollment in advanced classes suggest that the lack of achievement improvements arises in spite of what is generally viewed to be positive treatments. We also note that results for 6th grade

[34] We estimate this model such that each observation is assigned a weight equal to the teacher's share of classes taught to a student in a given subject. For example if a student takes a class in US history and another class in geography, then the student will have two observations in the social studies regression, one for each class, with a weight of ½ for each observation. Additionally, since the Stanford exams are given in January, we assign to each student the teachers they had in the spring of the previous academic year and the fall of the current academic year.

[35] Our model diverges from Kain and Staiger (2008) in two key ways. First, they use a random effects rather than fixed effects framework. We prefer the latter as it allows for weaker identification assumptions. Second, they utilize a Bayesian smoother that adjusts estimates for teachers with few observations towards the mean. While this strategy is important when trying to identify the influence of teachers on students, it is inappropriate in our context testing whether GT students receive higher quality teachers, as teachers with fewer observations will tend to be younger and less experienced, hence pushing their estimates to the mean would give us biased measures of actual teacher quality.

[36] Reduced form and first stage results are provided in Online Appendix Table 8.

are stronger as they show peer differences of 0.37 to 0.46 standard deviations as well as larger differences in Vanguard class enrollment. These results are provided in Online Appendix Table 9. Below we investigate potential explanations for these findings, but first we turn to our analysis of GT magnet lotteries.

IV. Estimates of the Impact of Attending a GT Magnet Using Randomized Lotteries

One reason the RD analysis does not show positive impacts from GT services on student outcomes may be that the qualification boundary is set low enough so that students who marginally qualify for GT services are not be able to take advantage of the purported benefits. Therefore, to examine other parts of the student quality distribution in this section we present results using lotteries for the two GT over-subscribed magnet middle schools. Because the lottery is random, the comparison is across the entire distribution of those who apply. In fact, not only are the lottery students stronger than the marginal GT students in the RD sample, they are stronger than the average GT student in the District as shown in Table 1. A disadvantage, however, is that the lottery losers have a range of alternative experiences, although the bulk of them are in neighborhood GT programs. Nonetheless, students in the magnet GT schools with lotteries are shown to receive a more intense experience than students in other GT programs.

1. Data

Our lottery sample is derived from the set of 5th grade students determined to be eligible for GT in 2007-08 who apply for admission to one of the two middle schools with an over-subscribed GT magnet program.[37] We restrict our analysis to students who are observed to be

[37] The application process involves a single form where students may apply to up to three of the eight magnet schools. Students also list which is their 1st, 2nd and 3rd choice schools. Unfortunately our data only informs us of

enrolled in LUSD in 5th 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 would be more likely to leave if they lose the lottery, as many have previously attended private or charter schools. In addition, we drop students zoned to one of the schools with a regular program for students in the attendance zone.[38]

While admission for non-zoned students is determined by a lottery, our data does not directly provide the lottery numbers or outcomes. Instead we identify whether a student is offered admission including those initially on a wait list. Students with an older sibling in the school are exempt from the lottery, but as discussed in Section III.2 above we believe the impact of this on our results is negligible. In total the sample includes 542 students who participate in a lottery. Of these 394 are offered admission and 148 are not. By 7th grade 440 students including 331 winners (84%) and 109 losers (74%) 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 non-compliance with the lotteries we employ a 2SLS strategy that instruments GT magnet attendance with lottery outcomes.[39]

---

whether a person is offered a spot or wait listed and does not have direct information on applications. Hence if a student is offered a spot at his or her first choice school we do not know if they applied for the other school. Nonetheless, we find no cases where a student is placed on the wait list for one of the lottery schools and offered a spot or waitlisted at the second while there are multiple instances whereby students are waitlisted at a lottery school and offered a slot at a non-lottery magnet. Hence, it appears that applying to both lottery magnets was very rare behavior in our data.

[38] The second school does not have zoned students, although it does include a program for students with severe physical disabilities such as blindness and deafness. Students who are enrolled in this alternative program are not included in our lottery sample.

[39] By 7th grade 67% of lottery winners attend a magnet with a lottery while 17% attend another school and 16% leave the district. For lottery losers, 18% attend a lottery campus in 7th grade while 56% attend a different school and 26% leave the district.

2. Tests of Validity of Lottery Design

Table 6 presents the balancing tests for the lottery sample. The results strongly suggest that the lotteries for both magnet middle schools are conducted in a random way, as the ex-ante baseline (5th grade) sample has no significant coefficient on any of the twenty covariates we test.[40] Further, using the ex-post estimation (7th grade) sample shows no significant differences between winners and losers except for math, which is significantly higher for winners at the 10% level. Although having one significant result out of twenty regressions 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 7th grade we find that lottery winners are 11 percentage points less likely to attrit (standard error of 0.04).

We thus use these 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 as well as demographic characteristics. Second, we use a weighting procedure in the regressions that mimics the original lottery sample in order to correct for potential attrition bias. To do this we reweight the sample by the inverse of the predicted probabilities from a probit of attrition on 5th grade student characteristics.[41] Third, we estimate bounds on the impact of GT using a procedure 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. Through a generalized method of moments (GMM) estimator, upper and lower bounds are then generated. The upper bound assumes students at risk of leaving due to losing the lottery have achievement equal to the

[40] Results for the 6th grade sample are similar as are results where standard errors are not clustered. These are provided in Online Appendix Tables 10 and 11.
[41] Results of the probit regression are provided in Online Appendix Table 12.

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.[42]

3. Results for GT Magnet Programs

Two-stage least squares estimates of the impact on student achievement from attending one of the two magnet GT programs are shown in Table 7. Reduced-form estimates are provided in Online Appendix Table 13.[43] We provide both unweighted (rows 1 and 2) and inverse probability weighted (rows 3 and 4) estimates where the latter corrects for possible attrition bias. In rows (5) and (6) we provide upper and lower bounds that account for potential attrition bias using the Engberg, et al. (2010) methodology.

The results in Table 7 using our preferred specification of weighting with controls (row 4) suggest that, with the exception of science, which shows a 0.28 sd improvement, there is little impact of attending a GT magnet on achievement or attendance.[44] Due to the small sample sizes the estimates are somewhat imprecise, particularly using the inverse-probability weighted model. Even so, we note that the point estimates in row (4) for math, reading and social studies are negative and the estimate for language is effectively zero.[45] Hence, we believe these estimates provide strong evidence of a lack of positive impact of attending a magnet on achievement other

[42] That is, the upper bound assumes students at risk of leaving have only average scores, while the lower bound assumes they are in the upper tail. These assumptions are those suggested by Engberg, et al. (2010).
[43] The first stage is always significant at the 1% level with point estimates of 0.57 (standard error of 0.06) for unweighted and 0.47 (0.11) for weighted regressions. Detailed first-stage results are available upon request.
[44] Note that teacher manipulation is not a concern in this identification strategy, hence we can use the attendance results with confidence. Additionally, we do not provide discipline results as only 4% of students in the lottery sample have any disciplinary infractions in 7th grade.
[45] Results for 6th grade, provided in Online Appendix Table 14, show somewhat larger, albeit still insignificant in the preferred model, impacts for math and language and no impact for science. They also show a significant negative impact on attendance of -0.6 percentage points (roughly one fewer day per year).

than in science.[46] The bounding analysis in rows (5) and (6) confirm the results in row (4).[47] Once again we see little to suggest that there is any substantial positive impact on math, reading, language and social studies. For science, the lower bound does drop to zero which suggests that the positive result there may be due to attrition bias, but it nonetheless confirms that there is at least no negative impact on science scores.[48]

In Table 8 we investigate to what extent there is an observable difference in treatment from attending a GT magnet. [49] The first five columns of the table show that using the weighted estimates, students who attend magnets gain peers, measured at the grade-course-teacher level, who score on average between 0.7 and 1.2 standard deviations higher than peers for the lottery losers.[50] Additionally, we find that students who attend a GT magnet gain teachers whose value-added estimates are 0.09, 0.03 and 0.04 sd higher in math, English and Social Studies, respectively.[51] Finally, in Online Appendix Table 18 we investigate whether there is any variation in the estimates by student types. Due to the small sample we are limited in how finely

---

[46] Top-coding of exams is a potentially even larger concern here than in the RD since the achievement levels of the lottery sample are higher. In Online Appendix Figures 14 - 18 we provide distribution plots of raw scores on 7th grade exams by lottery winners and losers. Although the mass of achievement is further to the right than in the RD sample, there nonetheless appears to be substantial room for achievement to improve for most students. Hence we do not believe that top-coding explains our lack of positive effects.

[47] We do not provide bounding analyses for attendance as it performs poorly when the mean outcome is centered near a top-code as it tends to estimate outcomes to be above the top-code, which is the case in this sample since mean 5th grade attendance rates are 98.0 with a maximum of 100.

[48] Online Appendix Table 15 shows lottery results when we use attending a lottery magnet specifically as treatment (e.g. place non-lottery magnets in same category as neighborhood GT) and when we identify students who are taken off the wait list as losing the lottery. In both cases the results are similar to baseline except that science impacts become statistically insignificant.

[49] Estimates without controls and estimates for 6th grade are provided in Online Appendix Table 16 and are similar to those shown in Table 9.

[50] In Online Appendix Table 17 we provide results under assumptions of student sorting into sections by ability or randomly as was done in Appendix Table 6 for the RD analysis. The results generally show similar levels of peer improvement where the difference does not fall below 0.5 standard deviations in any case.

[51] As in the RD analysis, we also look at differences in course level. While there are no significant differences in 7th grade in the likelihood of enrolling in Vanguard courses, in 6th grade students who attend magnets are approximately 10 percentage points more likely to take vanguard courses in math, English and social studies while they are 9 percentage points more likely to take Vanguard in science.

we can cut the data, but nonetheless we find no discernable patterns across subpopulations. Hence it is clear that GT magnet students gain large improvements in their educational environment yet experience little improvement in achievement except in science. In the next section we discuss some potential explanations for the lack of positive impacts in both the RD and lottery analyses despite the apparent improvements.

V. Discussion

Given that we have established that GT students experience substantial treatments including better peers, more advanced courses (in the RD analysis), and higher quality teachers (in the lottery analysis), it is perplexing that we find little evidence of positive impacts on achievement. One possibility explaining our findings is that our achievement measure is not well suited to discerning improvements in gifted students. This would be particularly worrisome if we were to use a state accountability exam targeted towards low achieving students, but less of an issue with the Stanford Achievement Test. Indeed, we have already shown that there is little evidence of bunching near the maximum score (top-coding) in either the RD or lottery samples. Nonetheless, it is possible that the additional course material taught in GT classes may not be well aligned with topics covered in the achievement test.[52] While we cannot rule out this possibility, we note that the improvement in peers would be expected to generate higher achievement even if the curriculum is not well targeted to the exam.

Another potential explanation is marginal students may suffer due to difficulty with more advanced material. In this view, the eligibility cut-off may be set at an inappropriate level as it

[52] Ideally, one would like to test some outcomes that might better align with the GT curriculum such as college-going, SAT scores, and AP/IB exam scores. Unfortunately, since the matrix and lottery data are only available for the last few years, not enough time has elapsed to investigate these outcomes.

leads the district to classify students who are unable to deal with the advanced GT material. While this explanation could be relevant for the RD results for marginal GT students, we demonstrate that the lottery sample includes higher achieving students for whom the advanced material would be more suitable.

Given the strength with which peer effects have been found to operate in several different contexts, one would expect that we would find achievement improvements simply from the peer effects alone (Angrist and Lang, 2004; Duflo, Dupas and Kremer, 2010; Hoxby and Weingarth, 2006; Imberman, Kugler and Sacerdote, forthcoming; Lavy, Paserman and Schlosser, 2008; Lavy and Schlosser, 2007). Nonetheless, one possible reason for finding no impact of the differential GT resources is that the peer effect, in addition to the potential benefits found in the literature cited above, has a potential cost as entering GT may reduce 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).

There is substantial evidence from the educational psychology and sociology literature, moreover, that students who are placed in higher achieving ability groups can be psychologically harmed. A measure that is commonly used in this literature is the idea of a student's "self concept," how a student perceives his or her own abilities relative to an objective metric such as achievement. Marsh, Chessor, Craven and Roche (1995) compare GT students to observably similar students in mixed GT and non-GT classes and find that GT students show declines in their math and reading self concept. Additionally, Zeidner and Schleyer (1999) find lower self concept and more test anxiety in gifted students in ability segregated classrooms. Similar results are also found by Preckel, Gotz and Frenzel (2010) and Ireson, Haliam and Plewis (2010). An alternative explanation is that teachers may target the material in their classes based on student ability (Duflo, Dupas, and Kremer, 2011).

Since we do not have direct evidence on student confidence, nor do we have direct evidence on how teachers target their classroom material, we empirically examine how the relative status changes for students in our two samples. Specifically, if student course grades and rank within their class changes based on their admittance to a GT program or to a selective school, it is possible that the conditions for invidious comparison exist.[53] Similarly, if course grades and rank change, it is possible that these changes may affect how well a student matches the ideal target student for a teacher. In Panel I of Table 9 we provide estimates of the impact of GT enrollment on course grades in the RD model, and of attending a GT magnet on grades in the lottery analysis. In both cases we find clear reductions in grades. For the RD sample grades fall by a statistically significant 4 points out of 100 (3 points changes a grade from a B+ to a B, for example) in math and by 2 to 3 points in other subjects, although these estimates are not statistically significant for 7th grade.[54] For the lottery analysis the grade reductions are even more dramatic with drops of 7 points in math, 8 in science, and 4 in social studies using the inverse-probability weighted regressions.

In addition to the raw grades it is useful to consider how students' rankings within their peer groups differ by treatment status, as this provides a direct measure of how a student may

[53] For example, although it is difficult to establish causality, some research has found a link between grades and students' self-concept (Marsh, Trautweing, Ludtke, Koller and Baumert, 2005), satisfaction (Howard and Maxwell, 1980), and self-worth/self-esteem (Crocker, Karpinski, Quinn and Chase, 2003; Owens, 1994).

[54] We do not show reading as more advanced students do not take reading in 7th grade. Nonetheless, in the RD sample we also find a significant drop in reading grades in 6th grade, and among students who take reading in 7th grade of four points. For the lottery sample, however, only a handful of students take reading in 7th grade and hence the estimates are too imprecise to draw inferences.

perceive his or her position in the achievement distribution. We assume that students mostly compare themselves to students who take the same courses in the same grade. Thus, we rank students within each school-grade-course cell by their final course grades and convert these rankings to percentiles. Figures 8 and 9 demonstrate that the rankings based on 7th grade courses exhibit notable drops when students cross the GT eligibility threshold. In panel II.A of Table 9, we provide regression results while adding controls for race, gender, economic disadvantage, LEP, and prior gifted status. The results show that marginal GT students have a relative rank 13 to 21 percentiles lower than marginal non-GT students in 7th grade. Panel II.B shows that attending a GT magnet in 7th grade generates a nearly 30 percentile ranking drop in all four of the courses examined.[55]

To the extent that the negative estimates for grades and rank in the RD analysis reflects absolute changes in learning, this suggests that the more difficult course work could be ill suited to students at the eligibility margin. We would not, however, expect the lottery participants to be ill suited to more difficult coursework given their high positions in the GT achievement distribution. Additionally, we find little difference in curricular differences as the likelihood of enrolling in a "Vanguard" class in 7th grade between lottery winners and losers is virtually identical. Thus, it seems likely that some portion of the grade effects reflect changes in relative rank independent of learning impacts. Whether these effects result from a student's self-perception, or because of how well the material presented by teachers matches the students' ability to learn is not clear.

[55] Note that we limit the 6th grade results to the 2007-08 cohort only in this analysis so that the sample is comparable to that used in the 7th grade achievement analysis.

VI. Summary and Conclusion

In this paper, we identify the impact of providing gifted and talented services on student achievement and behavior. We exploit a unique universal evaluation in a large urban school district in the Southwest US where all students are evaluated for GT eligibility in 5th grade regardless of prior GT status. This allows a regression discontinuity specification for students on either side of the eligibility cutoff, and we thus examine achievement, attendance and discipline differences by 7th grade. We also exploit a second data set, where two of the middle schools with GT magnet programs in this district are over-subscribed and conduct lotteries to determine admission amongst a pool of eligible GT students. Both of the samples offer a larger dispersion of student ability than is typical. In the RD dataset, the multiple criteria for eligibility results in a wide range of student achievement even for the marginal students. For the lottery sample, student applicants are on average very strong relative to even GT students in the district. The caveat in our strategies, however, is that the alternative to treatment varies. Marginal students not admitted to the GT program take "regular" classes, while students that lose the lottery receive GT services but in a less intensive atmosphere than that provided in the magnet schools.

Our analysis shows that both the RD and lottery samples meet the standard validity tests, with the main exception that lottery losers are more likely to leave the district. We correct for attrition in the lottery sample through inverse probability weighting for our estimates, and we also generate coefficient bounds using a procedure proposed by Engberg, et al. (2010).

The RD results indicate that GT services generate little impact on achievement for students on the margin of qualifying. For the lottery analysis we also find little evidence of improvement in achievement or attendance with the exception of science. These results are surprising given that we find large improvements in peer achievement on the order of 0.3 standard deviations in the RD analysis, and 0.7 to 1.2 standard deviations in the lottery analysis. In addition, we find that students on the margin of GT eligibility enroll in more advanced classes, while students that gain admission to the premier GT magnet schools gain higher quality teachers. The estimates from these two samples and specifications are reduced forms, in that they do not differentiate among the many mechanisms by which student achievement might be impacted. Nonetheless, we are able to rule out many of the standard explanations for the lack of observed improvement in achievement.

What we do find, however, is that that both raw course grades and students' relative rankings as measured by grades fall substantially in both the RD and lottery samples. These results establish that an invidious comparison model of peer effects may be operative, or that whether students match the target student for teachers is important. We therefore believe that our estimates show that how peer effects operate is rather subtle, and that the evidence we bring here suggests that how the relative standing of students changes within other institutional changes is important for discerning the total impacts.

VII. References

Abdulkadiroglu, Atila, Joshua D. Angrist and Parag A. Pathak, "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." NBER Working Paper No. 17264 (2011).

Angrist, Joshua D. and Kevin Lang, "Does Schooling Integration Generate Peer Effects? Evidence from Boston's Metco Program," *American Economic Review*, 94 (2004): 1613-1634.

Argys, Laura, Daniel Rees, and Dominic Brewer, "Detracking America's Schools: Equity at Zero Cost?" *Journal of Policy Analysis and Management* 15:4 (1996): 623-645.

Barecca, Alan I., Melanie Guldi, Jason M. Lindo, and Greg R. Waddell, "Running and Jumping Variables in Regression Discontinuity Designs." University of Oregon, mimeo (2010).

Betts, Julian, and Jaimie Shkolnik, "The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools." *Economics of Education Review* 19:1 (2000): 1-15.

Bhatt, Rachana, "A Review of Gifted and Talented Education in the U.S.," *Education Finance and* Policy, forthcoming, 2011.

---, "The Impacts of Gifted and Talented Education," mimeo, 2010.

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

Cicala, Steve and Roland Fryer, "A Roy Model of Social Interactions." NBER Working Paper No. 16880 (2011).

Clark, Damon, "Selective Schools and Academic Achievement," the B.E. Journal of Economic

Policy Analysis and Policy, 10, (Advances), Article 9.Crocker, Jennifer, Andrew Karpinski, Diane M. Quinn and Sara K. Chase, “When Grades Determine Self-Worth: Consequences of Contingent Self-Worth for Male and Female Engineering and Psychology Majors.” *Journal of Personality and Social Psychology* 85:3 (2003): 507 – 516.

Davis, James A., “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, July, 1966, 17-31.

Davis, Billie, John Engberg, Dennis N. Epple, Holger Sieg, and Ron Zimmer, “Evaluating the Gifted Program of an Urban School District using a Modified Regression Discontinuity Design,” NBER Working Paper #16414, September, 2010.

Dobbie, Will and Roland G. Fryer, “Exam High Schools and Academic Achievement: Evidence from New York City.” NBER Working Paper No. 17286 (2011).

Duflo, Esther, Pascaline Dupas, and Michael Kremer, “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” *American Economic Review*, 101:5 (2011): 1739-1774.

Engberg, John, Dennis Epple, Jason Imbrogno, Holger Sieg and Ron Zimmer, “Evaluating Education Programs That Have Lotteried Admission and Selective Attrition.” University of Pennsylvania, mimeo (2010).

Epple, Dennis, Elizabeth Newton, and Richard Romano, "Ability Tracking, School Competition, and the Distribution of Educational Benefits." *Journal of Public Economics* 83:1 (2002): 1-48.

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

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects With a Regression-Discontinuity Design." *Econometrica* 69:1 (2001): 201-209.

Hoekstra, Mark, "The Effect of Attending the Flagship University on Earnings: A Discontinuity-Based Approach." *Review of Economics and Statistics* 91:4 (2009): 717-724.

Howard, George S. and Scott E. Maxwell, "Correlation Between Student Satisfaction and Grades: A Case of Mistaken Causation?" *Journal of Educational Psychology* 72:6 (1980): 810 – 820.

Hoxby, Caroline and Gretchen Weingarth, "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects," American Economic Association Annual Meetings, http://www.aeaweb.org/annual_mtg_papers/2006/0108_1300_0803.pdf (2006).

Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote, "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review*, forthcoming.

Ireson, Judith, Susan Haliam and Ian Plewis, "Ability Grouping in Secondary Schools: Effects on Pupil's Self-Concepts." *British Journal of Educational Psychology* 71:2 (2010): 315–326.

Jackson, Kirabo, "Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago." *The Economic Journal*, 120:549 (2010): 1399-1429.Kane, Thomas J. and Douglas O. Staiger, "Estimating Teacher Impacts on

Student Achievement: An Experimental Evaluation," NBER Working Paper No. 14607 (2008).

Kulik, Chen-lin, and James Kulik, "Ability Grouping," in Handbook of Gifted Education, edited by Nicholas Colangelo and Gary Davis, 230-242. Boston: Allyn and Bacon (1997).

Lavy, Victor, M. Daniele Paserman and Analia Schlosser, "Inside the Black Box of Ability Peer Effects: Evidence from Variation in High and Low Achievers in the Classroom," NBER Working Paper no. 14415 (2008).

Lavy, Victor and Analia Schlosser, "Mechanisms and Impacts of Gender Peer Effects at School," NBER Working Paper no. 13292 (2007).

Lee, David, and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic* Literature, 2010, 48(2), 281-355.

Loveless, Tom, Steve Farkas and Ann Duffett, "High-Achieving Students in the Era of NCLB." Thomas B. Fordham Institute Report (2008).

Marsh, Herbert W., Danuta Chessor, Rhonda Craven and Lawrence Roche, "The Effects of Gifted and Talented Programs on Academic Self-Concept: The Big Fish Strikes Again." *American Educational Research Journal* 32:2 (1995): 285-319.

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142:2 (2008): 698 – 714.

Murphy, Patrick Ryan, "Essays on Gifted Education's Impact on Student Achievement." Doctoral dissertation, Florida State University (2009).

Neal, Derek, and Diane Whitmore Schanzenbach, "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *Review of Economics and Statistics* 92:2 (2010): 263-283.

Owens, Timothy J, “Two Diminsions of Self-Esteem: Reciprocal Effects of Positive Self-Worth and Self-Deprecation on Adolescent Problems.” *American Sociological Review* 59:3 (1994): 391-407.

Pop-Eleches, Cristian and Miguel Urquiola, “Going to a Better School: Effects and Behavioral Responses.” NBER Working Paper No. 16886 (2011).

Preckel, Franzis , Thomas Gotz and Anne Frenzel, “Ability Grouping of Gifted Students: Effects on Academic Self-Concept and Boredom.” *British Journal of Educational Psychology* 80 (2010):3: 451-472.

Reback, Randall, "Teaching to the Rating: School Accountability and the Distribution of Student Achievement," *Journal of Public Economics* 92:5-6 (2008): 1394-1415.

Rothstein, Jesse, "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *Quarterly Journal of Economics* 125(1), February 2010, p.p. 175-214.

Zeidner, Moshe and Esther Jane Schleyer, “The Big-Fish-Little-Pond Effect for Academic Self-Concept, Test Anxiety and School Grades in Gifted Children.” *Contemporary Educational Psychology* 24:4 (1999): 305 – 329.

Figure 1 - Gifted and Talented Matrix for GT Entry in 2008-09

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

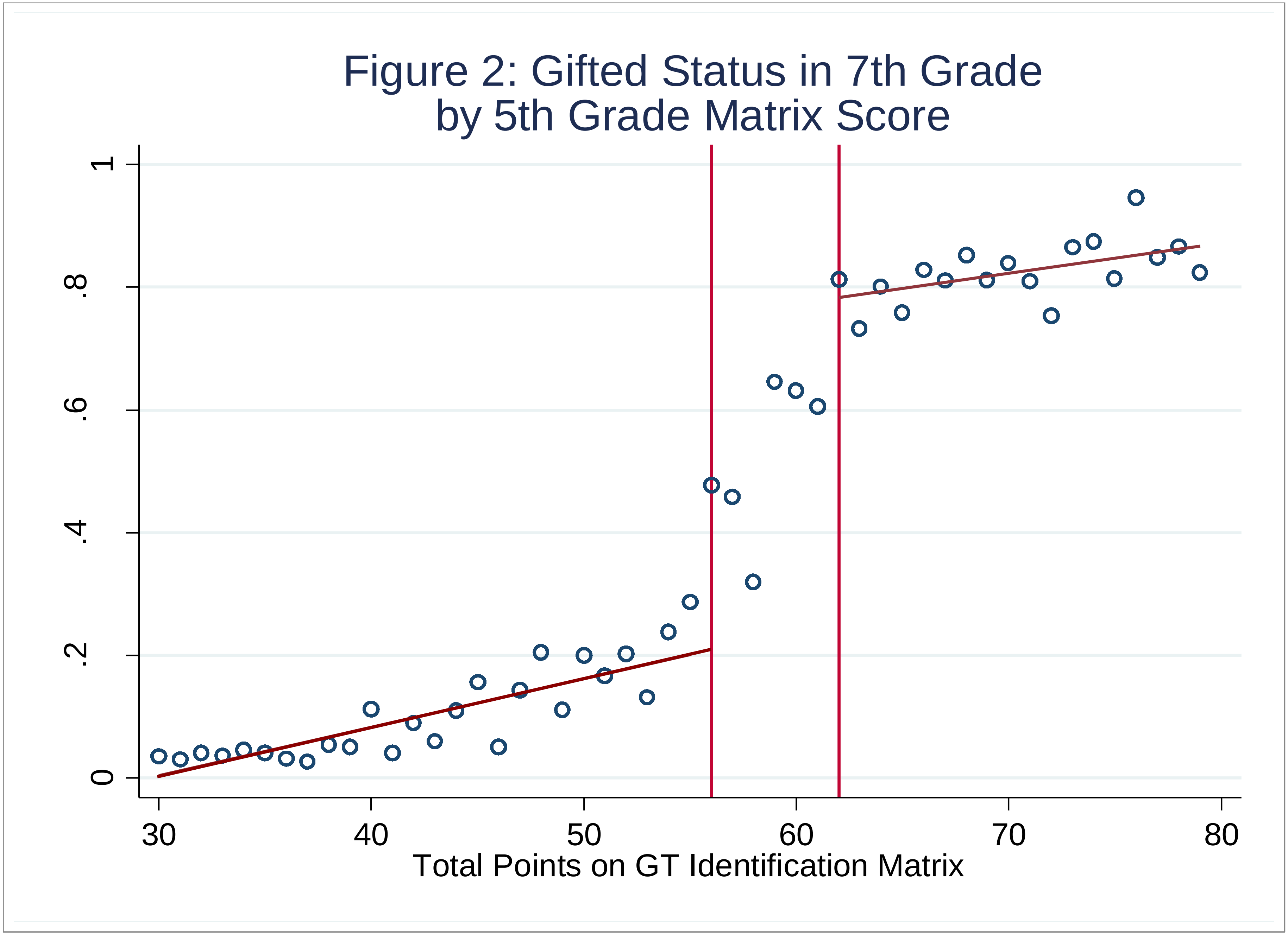
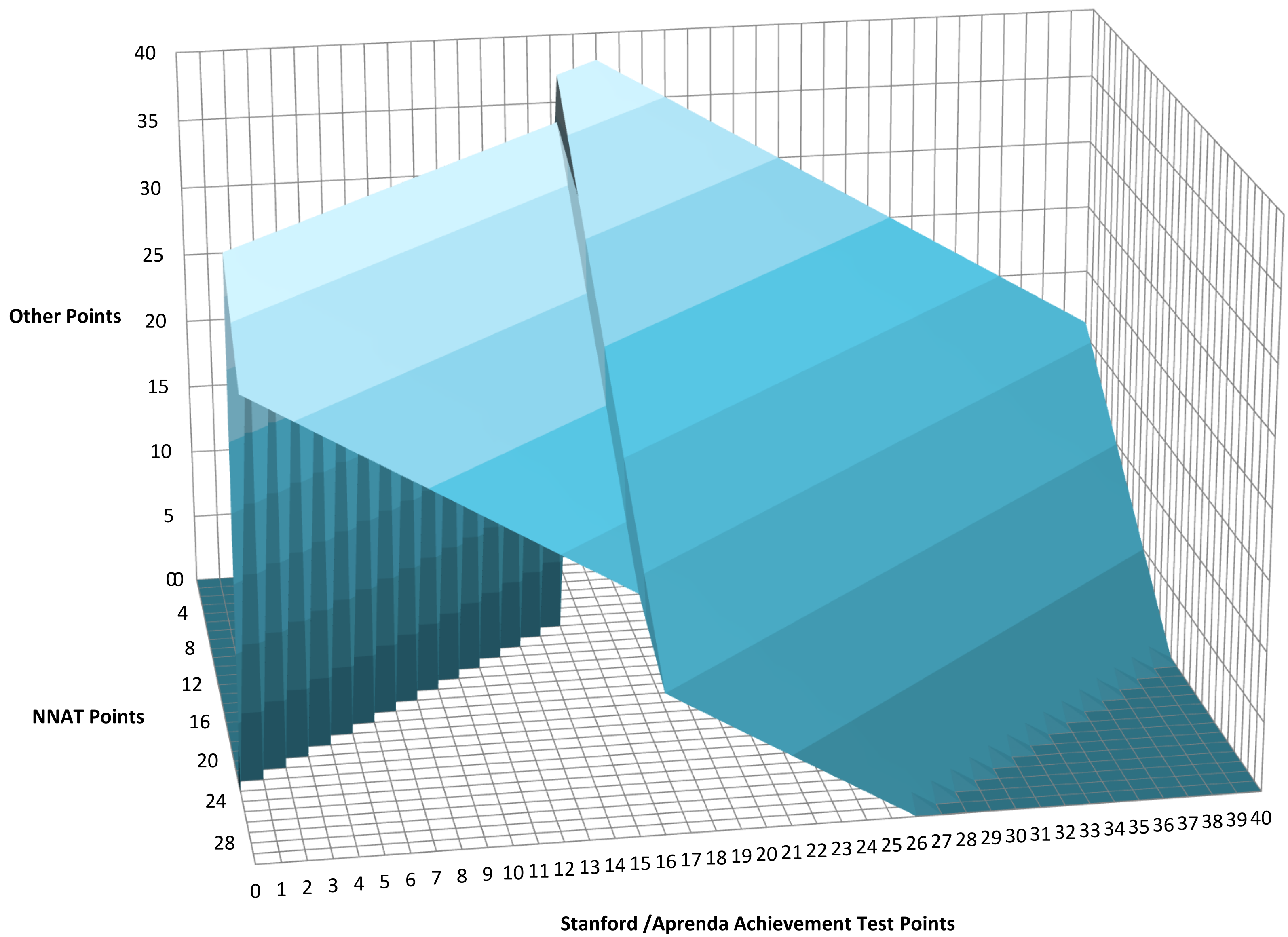


Figure 2: Gifted Status in 7th Grade
by 5th Grade Matrix Score
0
.2
.4
.6
.8
1
30
40
50
60
70
80
Total Points on GT Identification Matrix

Figure 3: Surface Plot of GT Qualification by Matrix Points



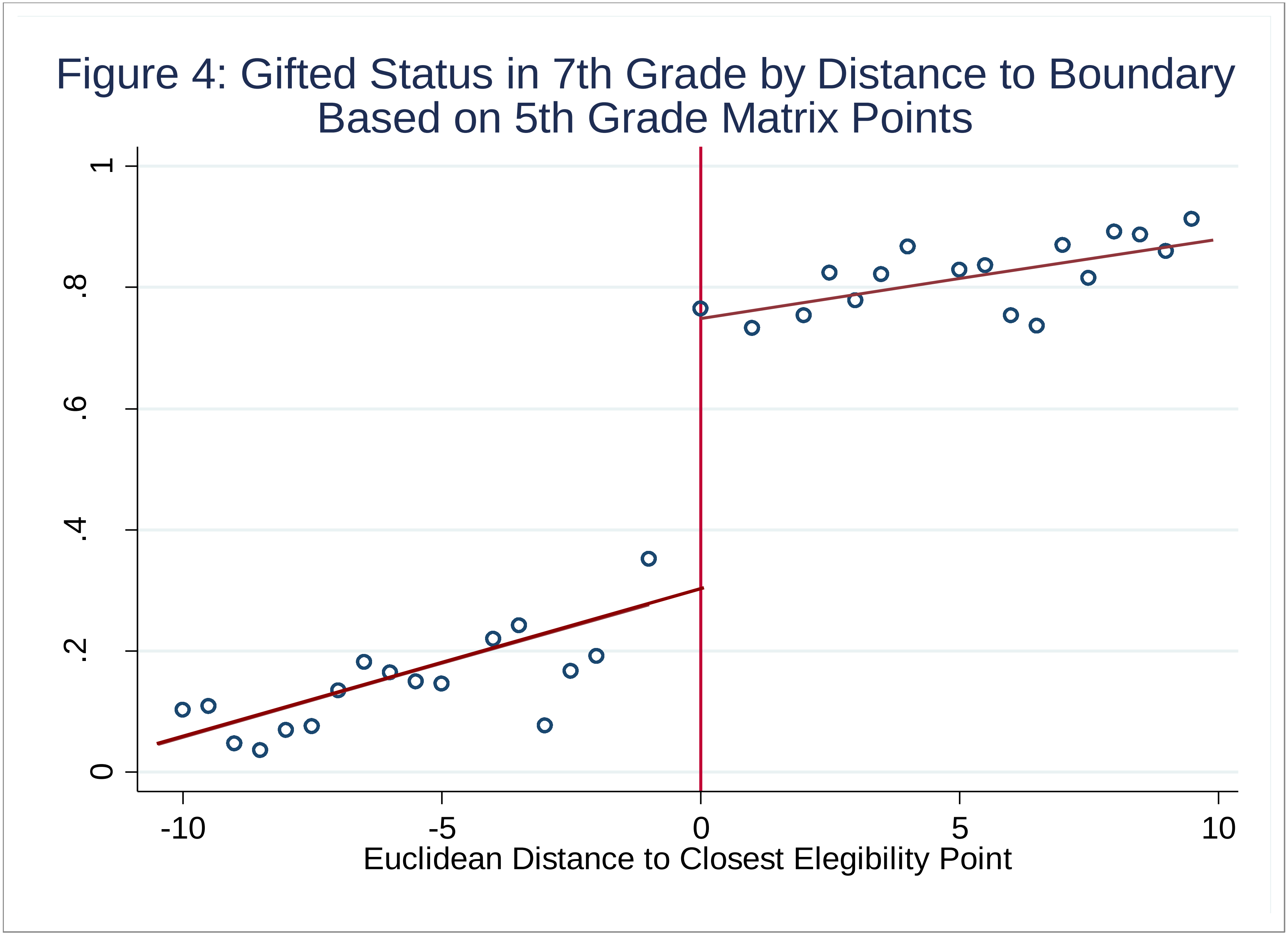


Figure 4: Gifted Status in 7th Grade by Distance to Boundary
Based on 5th Grade Matrix Points
0
.2
.4
.6
.8
1
-10
-5
0
5
10
Euclidean Distance to Closest Elegibility Point

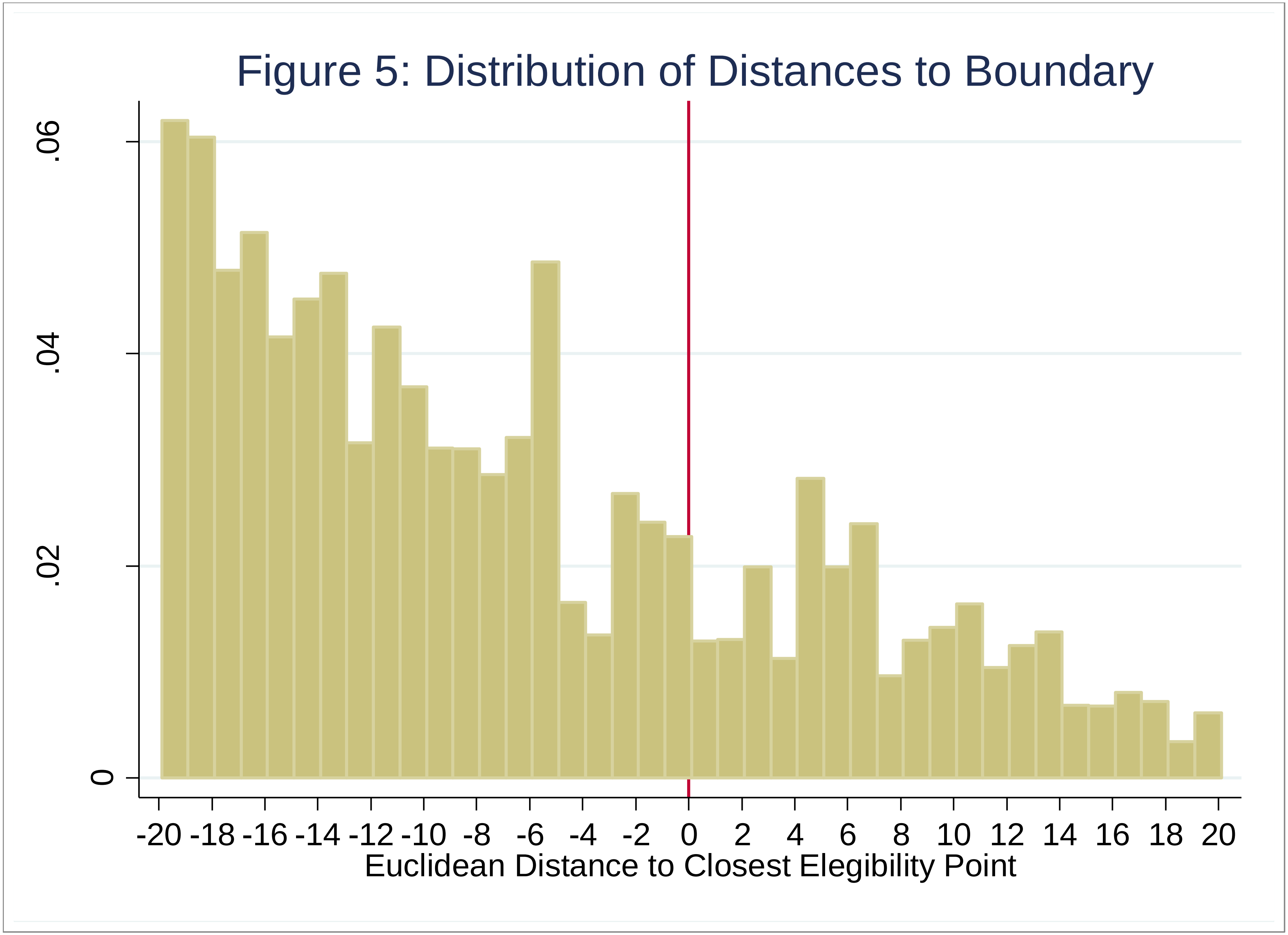


Figure 5: Distribution of Distances to Boundary
0
.02
.04
.06
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12 14 16 18 20
Euclidean Distance to Closest Elegibility Point

Figure 6: Stanford Math, Reading & Language in 7th Grade by Distance to Boundary

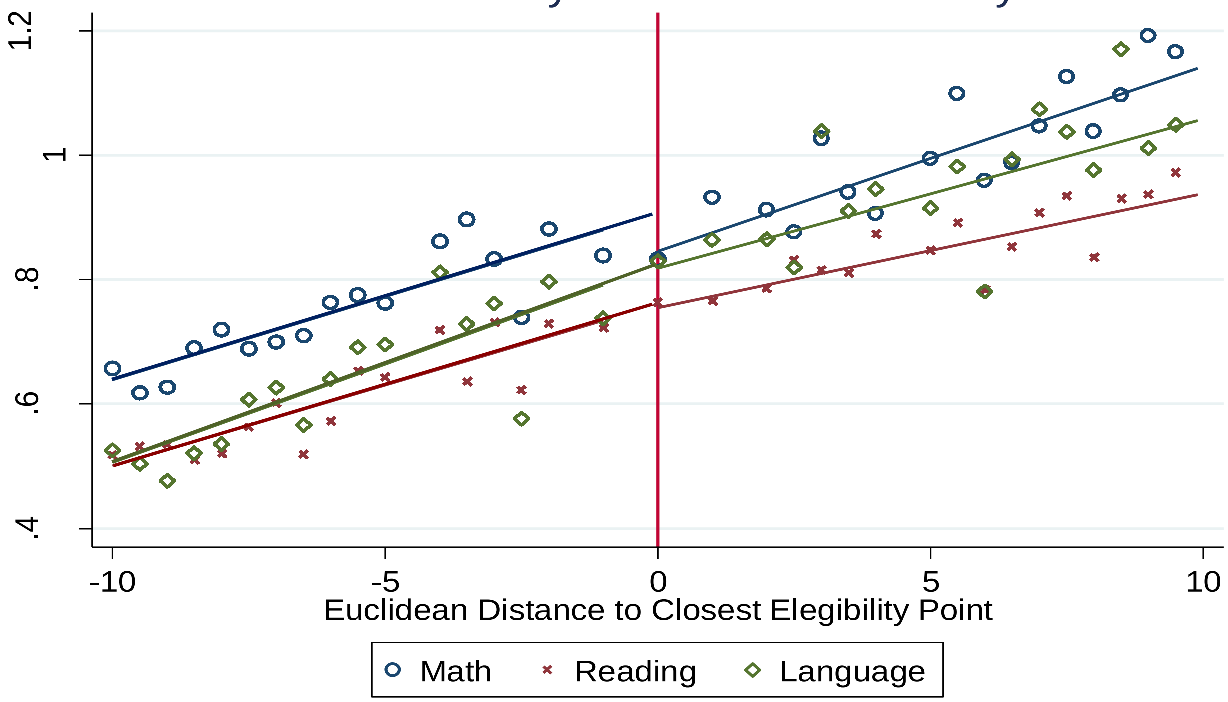


Figure 7: Stanford Social Studies & Science in 7th Grade by Distance to Boundary

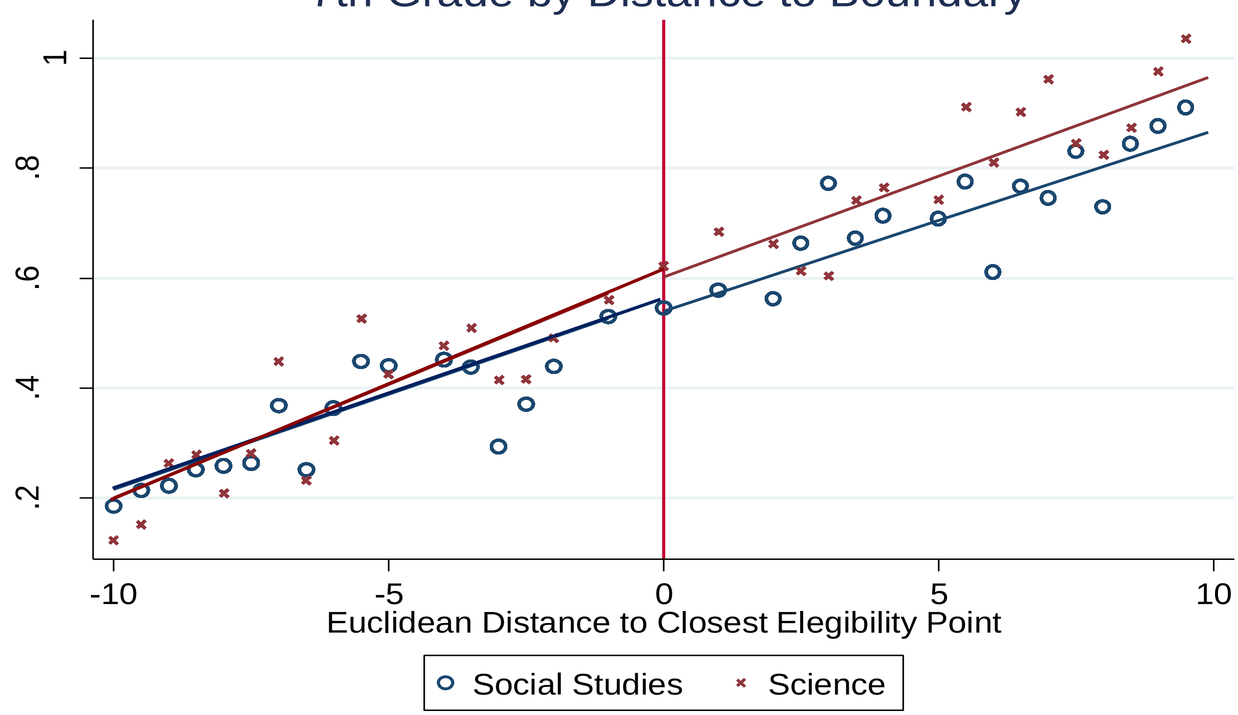
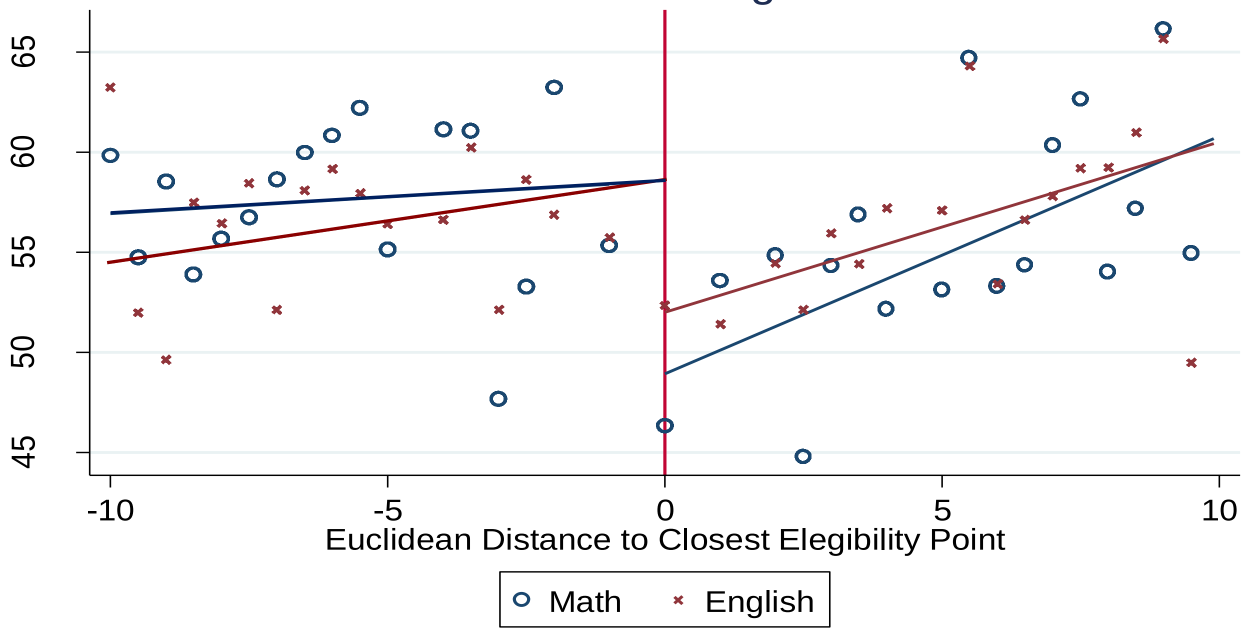
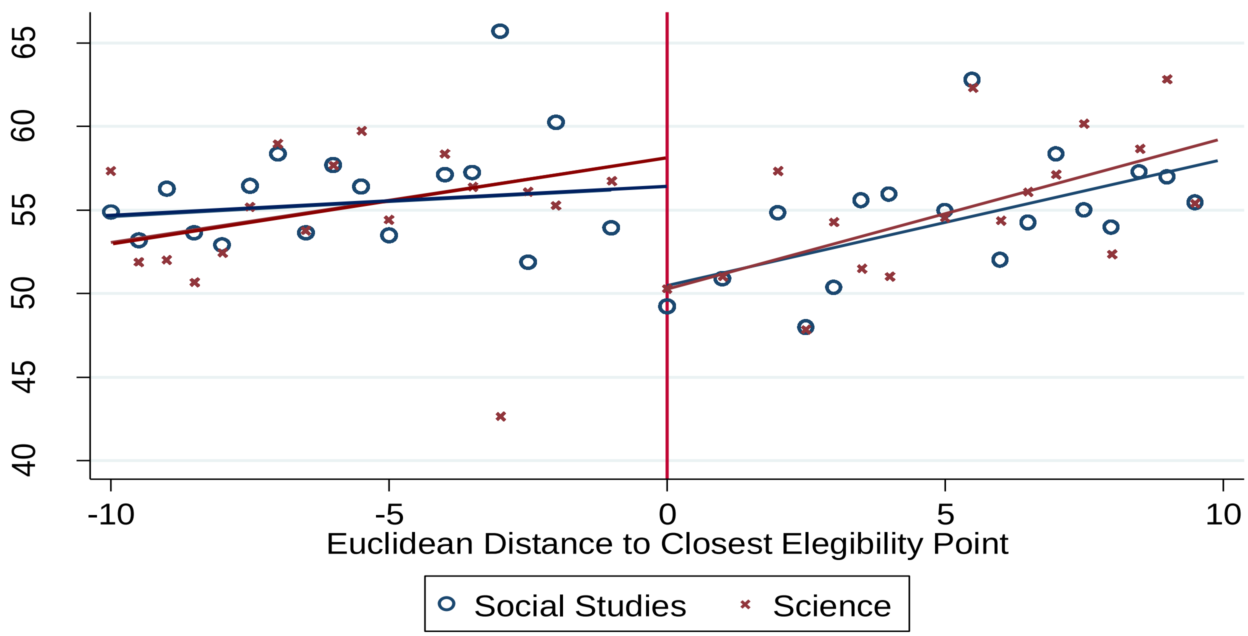


Figure 8: Rank in Course by Final Grade in 7th Grade by Distance to Boundary Math and English



Students with multiple courses in a subject are given the average rank over those courses.

Figure 9: Rank in Course by Final Grade in 7th Grade by Distance to Boundary Social Studies and Science



Students with multiple courses in a subject are given the average rank over those courses.

Table 1 - Characteristics of Students Evaluated for Middle School GT in 2007-08

	A. All 5th Grade Students			B. GT Magnet Lottery Sample		
	Gifted in 2009-10 (7th Grade)	Not Gifted in 2009-10	Not in Sample in 2009-10	In GT Magnet in 2009-10	Not in GT Magnet in 2009-10	Not in Sample in 2009-10
	A. 5th 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	0.74	0.06	0.18	1.61	1.39	1.72
	(0.59)	(0.39)	(0.47)	(0.79)	(0.71)	(1.03)
Stanford Reading	0.64	-0.02	0.11	1.72	1.60	1.83
	(0.41)	(0.39)	(0.47)	(0.78)	(0.77)	(0.87)
Stanford Language	0.74	-0.16	0.01	1.61	1.48	1.83
	(0.59)	(0.57)	(0.67)	(0.84)	(0.76)	(0.94)
Stanford Social Science	0.43	-0.61	-0.42	1.52	1.48	1.75
	(0.68)	(0.68)	(0.80)	(0.86)	(0.84)	(0.91)
Stanford Science	0.50	-0.50	-0.30	1.47	1.36	1.61
	(0.66)	(0.65)	(0.76)	(0.89)	(0.79)	(0.95)
Disciplinary Infractions	0.04	0.21	0.25	0.02	0.05	0.01
	(0.26)	(0.73)	(0.87)	(0.15)	(0.24)	(0.10)
Attendence Rate	98.26	97.25	96.58	98.35	97.98	97.00
	(2.35)	(4.52)	(4.95)	(2.00)	(2.34)	(3.75)
	B, 7th Grade Outcomes					
Stanford Math	1.11	-0.40	-	1.70	1.53	-
	(0.45)	(0.41)	-	(0.84)	(0.86)	-
Stanford Reading	0.95	-0.31	-	1.66	1.58	-
	(0.37)	(0.38)	-	(0.66)	(0.72)	-
Stanford Language	1.08	0.17	-	1.59	1.44	-
	(0.57)	(0.58)	-	(0.80)	(0.72)	-
Stanford Social Science	0.88	-0.09	-	1.70	1.51	-
	(0.64)	(0.60)	-	(0.88)	(0.80)	-
Stanford Science	1.00	-0.18	-	1.72	1.36	-
	(0.79)	(0.71)	-	(0.94)	(0.77)	-
Disciplinary Infractions	0.28	1.25	-	0.05	0.13	-
	(1.11)	(2.61)	-	(0.24)	(0.86)	-
Attendence Rate	97.37	95.02	-	97.84	97.57	-
	(3.19)	(6.13)	-	(2.52)	(3.16)	-
Observations	1,919	8,748	3,652	291	149	102

Standard deviations in parentheses. Achievement is measured in standard deviation units within grade and year across the district. Disciplinary infractions are the number of times a student is given a suspension or more severe punishment. Economically disadvantaged refers to students who qualify for free lunch, reduced-price lunch or another federal or state anti-poverty program.

Table 2 - Reduced-Form Estimates of Discontinuities in Pre-Existing (5th Grade) Student Characteristics

	Black	Hispanic	Female	LEP	Gifted in 5th Grade	Special Education	Free / Reduced-Price Lunch	Stanford - Math	Stanford - Reading	Stanford - Language
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Above GT Cutoff	-0.000	0.014	0.024	0.039	-0.050	0.005	0.049	-0.067***	0.006	0.006
	(0.029)	(0.038)	(0.042)	(0.040)	(0.047)	(0.011)	(0.037)	(0.026)	(0.026)	(0.041)
Observations	2,650	2,650	2,650	2,650	2,650	2,650	2,650	2,637	2,638	2,636

	Stanford - Social Studies	Stanford - Science	# of Disciplinary Infractions	Attendance Rate (%)	Any Missing Matrix Data	Teacher Score	Teacher Points	Enrolled	Enrolled (Free/ Reduced-Price Lunch)	Enrolled (Non-Free/ Reduced-Price Lunch)
	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Above GT Cutoff	0.040	0.004	-0.001	-0.269	0.000	2.965	0.497	0.049	0.054	0.039
	(0.049)	(0.042)	(0.028)	(0.190)	(0.008)	(2.715)	(0.321)	(0.030)	(0.037)	(0.053)
Observations	2,636	2,637	2,650	2,650	2,650	2,648	2,648	3,438	2,177	1,261

Achievement is measured in standard deviations of scale scores within grade and year. Disciplinary infractions are the number of infractions warranting a suspension or more severe punishment per 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 -10 and 10. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are robust to heteroskedasticity and clustered by 5th grade school. The estimation sample - students observed in LUSD two years after evaluation (7th grade) - is used is for columns (1) to (17) . Regressions using the full set of evaluated students provides similar results and is provided in the online appendix.

Table 3 - Regression Discontinuity Estimates of Impact of Receiving G&T Services

Model	Dependent Variable	Stanford Achievement Test: Math	Reading	Language	Social Studies	Science	Disciplinary Infractions	Attendance Rate (%)
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Baseline								
Reduced Form	Above GT Cutoff	-0.061**	-0.005	-0.004	-0.020	-0.011	-0.006	-0.691**
		(0.030)	(0.029)	(0.044)	(0.038)	(0.060)	(0.120)	(0.311)
2SLS - 1st Stage	Above GT Cutoff	0.440***	0.443***	0.442***	0.440***	0.440***	0.436***	0.438***
		(0.057)	(0.057)	(0.058)	(0.058)	(0.057)	(0.058)	(0.058)
2SLS - 2nd Stage	**Enrolled in GT**	**-0.138****	**-0.011**	**-0.008**	**-0.045**	**-0.025**	**-0.014**	**-1.578***
		(0.068)	**(0.065)**	**(0.100)**	**(0.085)**	**(0.135)**	**(0.276)**	**(0.802)**
Observations		2,612	2,614	2,612	2,610	2,612	2,653	2,652
B. With Individual Controls								
Reduced Form	Above GT Cutoff	-0.016	-0.001	0.005	-0.007	0.008	0.003	-0.502*
		(0.022)	(0.020)	(0.031)	(0.031)	(0.048)	(0.112)	(0.268)
2SLS - 1st Stage	Above GT Cutoff	0.465***	0.457***	0.457***	0.454***	0.456***	0.451***	0.456***
		(0.060)	(0.061)	(0.061)	(0.061)	(0.061)	(0.060)	(0.060)
2SLS - 2nd Stage	**Enrolled in GT**	**-0.035**	**-0.002**	**0.010**	**-0.016**	**0.017**	**0.007**	**-1.101***
		(0.047)	**(0.044)**	**(0.068)**	**(0.068)**	**(0.106)**	**(0.248)**	**(0.653)**
Observations		2,597	2,600	2,596	2,594	2,597	2,650	2,649
C. Using Synthetic Matrix Scores								
Reduced Form	Above GT Cutoff	-0.024	-0.028	-0.028	-0.054	0.002	0.088	0.346
		(0.028)	(0.020)	(0.039)	(0.041)	(0.059)	(0.130)	(0.309)
2SLS - 1st Stage	Above GT Cutoff	0.229***	0.232***	0.230***	0.228***	0.229***	0.230***	0.229***
		(0.038)	(0.038)	(0.039)	(0.039)	(0.038)	(0.038)	(0.038)
2SLS - 2nd Stage	**Enrolled in GT**	**-0.106**	**-0.121**	**-0.120**	**-0.236**	**0.011**	**0.382**	**1.509**
		(0.122)	**(0.085)**	**(0.170)**	**(0.188)**	**(0.256)**	**(0.568)**	**(1.328)**
Observations		2,579	2,580	2,579	2,576	2,578	2,619	2,618

Achievement is measured in standard deviations of scale scores within grade and year. Disciplinary infractions are the number of infractions warranting a suspension or more severe punishment per 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 10 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 (5th grade) dependent varable 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 -10 and 10. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are robust to heteroskedasticity and clustered by 7th grade school.

Table 4 - 2SLS Regression Discontinuity Estimates of Impact of Receiving G&T Services
Specification Checks

			Stanford Achievement Test						
		First Stage (1)	Math (2)	Reading (3)	Language (4)	Social Studies (5)	Science (6)	Disciplinary Infractions (7)	Attendance Rate (%) (8)
(1)	Quadratic Smoother	0.422***	0.120	0.007	0.246**	0.146	0.305*	-0.445	-0.565
		(0.064)	(0.112)	(0.071)	(0.111)	(0.135)	(0.159)	(0.505)	(1.253)
	Observations	2,609	2,597	2,600	2,596	2,594	2,597	2,650	2,649
(2)	Cubic Smoother	0.371***	0.057	-0.029	0.276	-0.019	0.409	-0.617	-0.455
		(0.103)	(0.238)	(0.157)	(0.203)	(0.244)	(0.332)	(0.745)	(2.036)
	Observations	2,609	2,597	2,600	2,596	2,594	2,597	2,650	2,649
(3)	Add Middle School Fixed Effects	0.460***	-0.014	0.007	0.041	0.009	0.023	0.067	-1.039*
		(0.057)	(0.037)	(0.041)	(0.065)	(0.065)	(0.112)	(0.249)	(0.600)
	Observations	2,609	2,597	2,600	2,596	2,594	2,597	2,650	2,649
(4)	Limited to Observations With	0.456***	-0.027	0.003	0.013	-0.004	0.029	0.068	-1.186*
	No Missing Matrix Data	(0.061)	(0.048)	(0.044)	(0.067)	(0.068)	(0.108)	(0.263)	(0.684)
	Observations	2,538	2,526	2,528	2,525	2,522	2,525	2,577	2,576
(5)	Limit to Students Who Have 16 or More Stanford	0.892***	-0.024	0.003	0.135	0.003	-0.049	-0.433	-1.057
	and 10 or More NNAT Points	(0.317)	(0.081)	(0.059)	(0.101)	(0.100)	(0.128)	(0.505)	(0.923)
	Observations	1,295	1,288	1,290	1,287	1,287	1,288	1,312	1,311
(6)	Limit to Students Who Less than 16 Stanford or 10	1.028**	0.042	-0.069	0.082	-0.011	0.046	-0.008	-1.064
	NNAT Points	(0.510)	(0.099)	(0.066)	(0.122)	(0.121)	(0.169)	(0.296)	(0.941)
	Observations	1,314	1,309	1,310	1,309	1,307	1,309	1,339	1,339
(7)	Distance Between -4 & 4	0.391***	0.116	-0.097	0.132	-0.029	0.338	-0.762	-0.835
		(0.085)	(0.167)	(0.111)	(0.159)	(0.170)	(0.246)	(0.518)	(1.647)
	Observations	849	845	848	845	842	844	860	859
(8)	Distance Between -8 & 8	0.462***	0.005	0.014	0.111	0.056	0.115	-0.162	-0.638
		(0.056)	(0.058)	(0.046)	(0.072)	(0.080)	(0.103)	(0.325)	(0.758)
	Observations	2,057	2,047	2,052	2,047	2,044	2,047	2,084	2,083
(9)	Distance Between -12 & 12	0.472***	-0.009	0.018	-0.013	0.007	0.019	0.001	-0.823
		(0.055)	(0.039)	(0.036)	(0.057)	(0.063)	(0.086)	(0.209)	(0.549)
	Observations	3,178	3,162	3,163	3,158	3,158	3,160	3,222	3,220
(10)	Distance Between -16 & 16	0.488***	-0.022	0.009	-0.015	-0.022	0.017	0.100	-0.438
		(0.055)	(0.035)	(0.030)	(0.045)	(0.061)	(0.077)	(0.179)	(0.497)
	Observations	3,756	3,735	3,736	3,731	3,729	3,733	3,806	3,804
(11)	Local Linear Regressions	-	0.073	0.000	0.019	0.056	0.222	1.476	-0.434
	with Rectangular Kernel	-	(0.117)	(0.072)	(0.186)	(0.080)	(0.177)	(1.002)	(1.203)
	Observations	-	1,075	1,078	708	2,044	1,074	429	1,092
	Bandwidth for LLR (from Leave-One-Out Cross Validation)	-	5	5	3	8	5	2	5

Achievement is measured in standard deviations of scale scores within grade and year. Disciplinary infractions are the number of infractions warranting a suspension or more severe punishment per year. Controls for race, gender, economic disadvantage, LEP, prior gifted status and lagged (5th grade) dependent varable 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 -10 and 10. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are robust to heteroskedasticity and clustered by 7th grade school.

Table 5 - 2SLS Estimates of Impacts of G&T Services
Effects on Educational Environment and Student Choices

	Peer Math Scores in Math Classes (1)	Peer Reading Scores in Read/Eng Classes (2)	Peer Lang Scores in Read/Eng Classes (3)	Peer Soc Scores in Soc Classes (4)	Peer Science Scores in Science Classes (5)	# of Core Regular Classes (6)	# of Core Vanguard Classes (7)	Enrolled in Vanguard Math (8)	Enrolled in Vanguard English (9)
Enrolled in GT	0.348**	0.287*	0.311**	0.235*	0.272*	-0.014	1.145*	0.315*	0.241
	(0.166)	(0.156)	(0.146)	(0.132)	(0.150)	(0.267)	(0.624)	(0.158)	(0.171)
Observations	2,629	2,494	2,494	2,567	2,567	2,643	2,643	2,629	2,497

	Enrolled in Vanguard Social Science (10)	Enrolled in Vanguard Science (11)	Attends Zoned School (12)	Attends Non-Zoned GT Magnet Campus (13)	Attends Other Non-Zoned (14)	Math Teacher Fixed Effect (15)	Read/Eng Teacher Fixed Effect (16)	Science Teacher Fixed Effect (17)	Social Science Teacher Fixed Effect (18)
Enrolled in GT	0.282*	0.282*	-0.050	0.260**	-0.210**	-0.001	0.016	0.005	0.014
	(0.165)	(0.165)	(0.109)	(0.109)	(0.098)	(0.025)	(0.010)	(0.014)	(0.013)
Observations	2,567	2,567	2,623	2,623	2,623	2,650	2,621	2,621	2,621

Achievement is measured in standard deviations of scale scores within grade and year. Teacher fixed effects are estimates from a student-level regression of achievement on lagged achievement, peer lagged achievement, race, gender, special education, LEP, at-risk status, teacher fixed-effects and school fixed-effects. Controls for race, gender, economic disadvantage, LEP, prior gifted status and lagged (5th grade) dependent varable included. Also includes a linear smoother with a slope shift above the cutoff. Peers are defined by teacher-course id-grade cells. The sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between -10 and 10. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are robust to heteroskedasticity and clustered by 7th grade school.

Table 6 - Balancing Tests for GT Magnet Lotteries - Covariates Measured in 5th Grade

Sample	Asian (1)	Black (2)	Hispanic (3)	White (4)	Econ Disadv (5)	Female (6)	At-Risk (7)	Special Education (8)	LEP (9)	Gifted (10)
Ex-Ante - Baseline Lottery	-0.030 (0.044)	0.030 (0.038)	0.041 (0.044)	-0.041 (0.050)	-0.035 (0.045)	-0.006 (0.047)	-0.011 (0.010)	-0.019 (0.017)	-0.033 (0.022)	-0.028 (0.035)
Observations	542	542	542	542	542	542	542	542	542	542
Ex-Post - Estimation Sample	-0.027 (0.048)	0.041 (0.038)	0.042 (0.055)	-0.057 (0.056)	-0.050 (0.059)	-0.001 (0.052)	-0.009 (0.011)	-0.015 (0.023)	-0.031 (0.027)	-0.024 (0.047)
Observations	437	437	437	437	437	437	437	437	437	437

			Stanford Achievement Test							
Sample	GT Magnet (11)	Total Matrix Points (12)	Math (13)	Reading (14)	Language (15)	Social Studies (16)	Science (17)	Attendance Rate (18)	Infractions (19)	Teacher Score (20)
Ex-Ante - Baseline Lottery	0.035 (0.030)	0.243 (0.926)	0.027 (0.069)	0.073 (0.063)	-0.034 (0.077)	0.053 (0.089)	0.010 (0.076)	-0.180 (0.201)	-0.022 (0.021)	0.029 (1.304)
Observations	542	542	540	541	539	540	539	542	542	536
Ex-Post - Estimation Sample	0.055 (0.045)	0.909 (1.173)	0.128* (0.074)	0.100 (0.075)	-0.059 (0.077)	0.063 (0.096)	0.090 (0.088)	-0.064 (0.230)	-0.022 (0.025)	-1.005 (1.471)
Observations	437	437	437	437	436	437	436	437	437	434

Achievement is measured in standard deviations of scale scores within grade and year. Disciplinary infractions are the number of infractions warranting a suspension or more severe punishment per year. Lotteries for two schools were conducted in 2007-08 hence regresions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student won the lottery. Robust standard errors clustered by 5th grade school in parentheses. Results without clustering are similar and provided in the online appendix.

Table 7 - Effect of Attending a GT Magnet School Relative to a GT Neighborhood Program

	Model	Stanford Achievement Test Math (1)	Reading (2)	Language (3)	Social Studies (4)	Science (5)	Attendence Rate (%) (6)
(1)	2SLS - Unweighted, No Controls	0.042	0.023	0.102	0.039	0.249**	-0.434
		(0.178)	(0.103)	(0.065)	(0.083)	(0.114)	(0.636)
	Observations	437	438	436	437	437	440
(2)	2SLS - Unweighted, Controls	-0.100	-0.058	0.142*	-0.032	0.208*	-0.425
		(0.112)	(0.105)	(0.081)	(0.098)	(0.119)	(0.411)
	Observations	437	438	435	437	436	440
(3)	2SLS - Weighted, No Controls	-0.266	-0.130	-0.060	-0.120	0.243	0.043
		(0.291)	(0.221)	(0.148)	(0.214)	(0.201)	(1.996)
	Observations	436	437	435	436	436	439
(4)	2SLS - Weighted, Controls	-0.224	-0.018	0.001	-0.036	0.281**	0.364
		(0.171)	(0.172)	(0.114)	(0.136)	(0.130)	(1.489)
	Observations	436	437	435	436	436	439
(5)	Engberg, Epple, Imbrogno, Sieg, Zimmer	-0.019	-0.095	0.074	-0.064	0.344*	-
	(2011) Bounds - Upper Bound	(0.196)	(0.157)	(0.162)	(0.185)	(0.180)	-
	Observations	437	438	436	437	437	-
(6)	Engberg, Epple, Imbrogno, Sieg, Zimmer	-0.353	-0.310	-0.207	-0.389	-0.013	-
	(2011) Bounds - Lower Bound	(0.251)	(0.192)	(0.215)	(0.249)	(0.248)	-
	Observations	437	438	436	437	437	-

Achievement is measured in standard deviations of scale scores within grade and year. Lotteries for two schools were conducted in 2007-08 hence all regresions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student is enrolled in a GT magnet program in 7th grade. Robust standard errors clustered by 7th grade school in parentheses. Results without clustering are similar and provided in the online appendix. Controls include indicators during 5th 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 for the estimator to converge. Finally, we do not provide bounds for attendance due to poor performance with censored data. See paper for details.

Table 8 - Treatments from Attending a GT Magnet School Relative to a GT Neighborhood Program

	Mean Peer Achievement (Std Deviations)					Teacher Fixed Effects			
	Math in Math Class	Reading in English Class	Language in English Class	Social Studies in Soc Class	Science in Science Class	Math	English/ Reading	Social Studies	Science
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2SLS - Unweighted, Controls	1.066***	0.659***	0.579***	0.794***	0.524***	0.081***	0.032**	0.031*	0.017
	(0.145)	(0.149)	(0.120)	(0.123)	(0.122)	(0.015)	(0.013)	(0.017)	(0.014)
Observations	440	436	436	439	439	440	440	440	440
2SLS - Weighted, Controls	1.164***	0.751***	0.686***	0.952***	0.659***	0.085***	0.032***	0.041**	0.016
	(0.179)	(0.172)	(0.143)	(0.180)	(0.166)	(0.019)	(0.011)	(0.019)	(0.013)
Observations	439	435	435	438	438	439	439	439	439

Achievement is measured in standard deviations of scale scores within grade and year. Teacher fixed effects are estimates from a student-level regression of achievement on lagged achievement, peer lagged achievement, race, gender, special education, LEP, at-risk status, teacher fixed-effects and school fixed-effects. Lotteries for two schools were conducted in 2007-08 hence all regresions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student is enrolled in a GT magnet program in 7th grade. Peers are defined by teacher-course id-grade cells. Robust standard errors clustered by 7th 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 5th 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. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 9 - 2SLS Estimates of Impacts of G&T on Course Grades and Rank (2007-08 Evaluation Cohort)

	I. Course Grades				II. Rank in Course (Percentiles)			
	Math (1)	English (2)	Social Studies (3)	Science (4)	Math (5)	English (6)	Social Studies (7)	Science (8)
A. Regression Discontinuity Analysis								
i. 7th Grade								
Enrolled in GT	-4.142**	-2.621	-2.473	-1.501	-21.1***	-15.5**	-17.1***	-13.2**
	(1.616)	(1.744)	(1.645)	(1.052)	(6.9)	(7.1)	(5.8)	(6.0)
Observations	2,643	2,510	2,581	2,602	2,643	2,510	2,581	2,602
ii. 6th Grade								
Enrolled in GT	-3.422***	-1.953	-2.931**	-3.411**	-17.9***	-16.9***	-22.6***	-22.9***
	(1.179)	(1.491)	(1.355)	(1.442)	(6.2)	(6.5)	(7.0)	(6.9)
Observations	2,739	2,609	2,754	2,733	2,739	2,609	2,754	2,733
B. Lottery Analysis (7th Grade)								
Unweighted, Controls	-8.283***	-4.096**	-4.062**	-6.988***	-29.5***	-27.1***	-27.8***	-29.3***
	(1.660)	(1.561)	(1.654)	(1.309)	(4.8)	(5.0)	(6.5)	(6.3)
Observations	440	437	439	439	440	437	439	439
Weighted, Controls	-7.311***	-2.719	-4.733**	-8.121***	-30.7***	-30.4***	-33.8***	-36.1***
	(1.847)	(1.990)	(1.733)	(2.297)	(5.4)	(7.7)	(6.8)	(8.9)
Observations	439	436	438	438	439	436	438	438

Rank is determined by rank-ordering the final grade in each course within school, grade and year, converted to percentiles. RD: Controls for race, gender, economic disadvantage, LEP, and prior gifted status are included along with 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 -10 and 10. Standard errors are robust to heteroskedasticity and clustered by 7th grade school. Lottery: Lotteries for two schools were conducted in 2007-08 hence all regresions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student is enrolled in a GT magnet program in 7th grade. Peers are defined by teacher-course id-grade cells. Robust standard errors clustered by 7th 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 5th grade for race, gender, special education, LEP, at-risk status, gifted, and whether the student was enrolled in a GT magnet. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.