Does Gifted Education Work? For Which Students?

David Card
UC Berkeley
and NBER

Laura Giuliano
University of Miami

March 2014
(Preliminary/Incomplete version)

ABSTRACT

Education policy makers have struggled for decades with the question of how to best serve high ability "gifted" students. A key issue of contention is whether eligibility for gifted programs should be based mainly on IQ, or on broader measures that take better account of both cognitive *and* non-cognitive skills. Using data from a large urban school district, we study the impacts of an intensive gifted education program that provides the same treatment to three distinct groups of fourth grade students: non-disadvantaged students with IQ scores of 130 or more; subsidized lunch participants and English language learners with IQ scores of 116 or more; and high-achieving students who do not meet the above IQ cutoffs, but qualify through high scores on state achievement tests. Regression discontinuity (RD) estimates based on the IQ thresholds for the first two groups show no effects on reading or math achievement. In contrast, RD estimates based on test score ranks for the high-achieving group show significant gains in reading and math, with treatment-on-the treated effects of 0.2 to 0.3 standard deviation units. Our results suggest that programs for high-potential students may be more effective for students selected on the basis of achievement than for those selected on the basis of IQ alone.

[*] We are extremely grateful to Cynthia Park and Jacalyn Schulman for their assistance in accessing and interpreting the data used in this study, and to Sydnee Caldwell, Yosub Jung, Attila Lindner, and Carl Nadler for outstanding research assistance. We also thank Kelly Bedard, Jesse Rothstein, Enrico Moretti, Chris Walters, and seminar participants at UC Berkeley, UC Riverside, UC Santa Barbara, UMass Amherst and the NBER Summer Institute for helpful discussions and suggestions. This research was supported by a grant from the Institute of Education Sciences (R305E110019).

Over the past century gifted education programs for high-ability students have expanded from a handful of cities to serve nearly seven percent of the U.S. student population.[1] A critical issue in the design of these programs is targeting. Early gifted programs often used high-IQ as the single criterion for eligibility. But by the 1970s, this practice was under attack: critics charged that IQ tests were biased against disadvantaged groups who were typically underrepresented, and argued more broadly that eligibility should be based on an array of both cognitive *and* non-cognitive traits.[2] The continuing debate is reflected in current state law. Most states still emphasize IQ as a criterion and more than one third mandate the use of IQ score cutoffs. However, the majority of states now have flexible policies for disadvantaged students, and the majority identify gifted students using a matrix of cognitive and non-cognitive measures (McClain and Pfeiffer, 2012).

Despite the divergent views of specialists and policymakers, there is almost no credible evidence on the relative effectiveness of gifted education programs for different target groups. In fact, there is little evidence even on the *average* impact of these programs; the best study to date (Bui et al., 2011) suggests gifted programs have little overall effect on standardized test scores.[3] In this paper, we use detailed administrative data from one of the country's largest school districts ("the District") to study the impacts of an intensive gifted education program for fourth and fifth grade students. Further, the design of the district's gifted programs allows us to compare the impact of the same treatment on three different target groups.

---

[1] Chapman (1988), Jolly (2009) and VanTassel-Baska (2010) discuss the history of gifted education in the U.S. Major milestones include Title I of the National Defense of Education Act (1958), which introduced federal funding for gifted education, and the Javits Act of 1988 which provides funding and support for research on gifted education.

[2] See Renzulli (1978) for example. The Marland Report to the U.S. Congress (Marland, 1972) introduced an expanded definition of gifted status that included children "with demonstrated achievement and/or potential ability" in one of several areas.

[3] There is a large related literature on ability tracking that is relatively inconclusive about the benefit of separate classes for higher-ability students (e.g., Slavin, 1987; Figlio and Page, 2002), though many studies lack rigorous research designs (Betts, 2011). Exceptions are Duflo, Dupas and Kremer (2010), Jackson (2010), Pop-Eleches and Urquiola (2013), and Vardadottir (2013). An extreme form of tracking is exam-based schools, which are studied by Abdulkadiroglu et al. (2011) and Dobbie and Fryer (2011).

The rules that govern eligibility for the district's gifted program create three distinct groups of participants who are placed together in separate gifted classrooms. First, there are non-disadvantaged students who meet a 130 cutoff on a standard IQ test, which is the crucial requirement for gifted eligibility (the District calls these "Plan A" gifted students). Second, there are students who meet a lower 116-point cutoff set for free- and reduced-price lunch participants and English-language learners (known as "Plan B" gifted students).[4] Third, any remaining seats in the gifted classrooms (typically more than one-half the slots) are filled by **non-gifted** students in the school/grade cohort who scored highest in statewide achievement tests in the previous year (these are known as "high-achievers").

Because students from all three groups are present in a typical gifted classroom, we have the unique opportunity to compare the effectiveness of the same program for regular and disadvantaged students selected on the basis of cognitive test scores (the Plan A and Plan B groups) versus those selected on the basis of past achievement (the high achievers). Moreover, the eligibility rules enable us to implement regression discontinuity (RD) methods that provide credible estimates of the program impacts for the three groups.

One issue for our IQ-based RD design is that test administrators appear to boost the scores of some marginally eligible students to meet the thresholds.[5] For higher-income Plan A students and to a lesser extent for Plan B students, we find many more IQ scores just above the cutoff threshold than just below. Nevertheless, for both Plan A and Plan B students, the third grade test scores of students on either side of the threshold are very similar, implying that conventional RD models may yield unbiased program estimates. We compare the results from these specifications to estimates from "donut-hole" specifications that ignore data close to the threshold (Barreca et al., 2010; Bajari et al., 2012), and first differenced specifications that are robust to IQ manipulation.

[4] The 130 cutoff is two standard deviations about the mean, and the 116 cutoff is roughly one standard deviation above the mean.

[5] As explained below we use students' first recorded IQ tests to avoid selective re-testing, and limit our sample to students whose first test was conducted by a District psychologist to avoid selective reporting of initial scores.

For Plan A gifted students, none of the models show positive impacts on achievement in math, reading or writing. For the disadvantaged Plan B gifted students, the models again all show zero impacts on math and reading achievement. However, there is some evidence of a positive impact on writing scores.

In contrast to the negligible impacts on reading and math scores for the Plan A and Plan B groups, RD models based on lagged achievement score ranks show large positive effects for non-gifted high achievers who are placed in gifted classrooms. We confirm this conclusion using an alternative research design that compares the top 20 non-gifted students at schools where there were no gifted children in the cohort (and no gifted class for fourth grade) to students at schools where there were 1-4 gifted students (and hence a gifted class with about 20 seats for high achievers). This design also allows us to verify that moving the highest achievers to a separate classroom has no effect on the scores of the next-highest group of students (ranked 25-44 in their cohort) – a spillover that could bias our RD design.

Our conclusion that the District's gifted program has little or no impact on the reading and math scores of gifted students is consistent with the results in Bui et al. (2011) and with the findings in recent studies of selective admission schools (Abdulkadiroglu et al., 2014; Dobbie and Fryer 2011). More surprising is our conclusion that placement in a gifted classroom has relatively large positive effects for "non-gifted" high achievers. The contrast in results between high achievers and Plan B gifted students is particularly compelling, because the two groups have similar prior achievement scores and because they tend to be concentrated in the same lower-income schools. Indeed, while we find positive effects across all schools, the effects are especially large in lower-income schools and among the disadvantaged high-achievers in those schools.

Our findings have two main policy implications. First, our results suggest programs for high-potential students may be more effective for students selected more on the basis of past achievement than on the basis of IQ. In fact, we will see below that because of the local nature of the estimates in our fuzzy RD design, our findings show the benefits of a gifted classroom environment are systematically larger for "over-achievers" (children with high achievement scores but unremarkable IQ scores) than for the "under-achievers" (high-IQ students with lower achievement scores) who are a traditional concern of many gifted programs. This suggests the prior achievement scores of high-achievers may signal they possess non-cognitive traits (e.g., high-motivation) that allow them to take advantage of programs that group capable students together in the same classroom.

The second implication is for high-achievers in lower-income schools. One effect of the District's Plan B policy for disadvantaged students has been to establish gifted programs and separate gifted classrooms in many lower-income schools where no gifted students would have been identified under the higher "Plan A" IQ cutoff. However, even under the Plan B cutoff very few gifted students are identified in lower-income schools. Hence the District fills up the gifted classrooms in these schools with high-achievers—typically at least three-quarters of the students in these classrooms are non-gifted high-achievers. Thus the "Plan B" policy also in effect establishes a gifted program in lower-income schools for high-achievers who otherwise would have been excluded by the IQ cutoffs. Importantly, this de facto tracking program for high achievers in lower-income schools has produced substantial effects.

II. Gifted Education in the District

Most elementary schools in the District provide part-time individualized instruction in first through third grades for students who have been identified as gifted. Since 2004 the District has required schools to offer separate gifted classrooms in fourth and fifth grades for all gifted students, with any open seats allocated to students who performed best in the previous year's standardized tests. Few schools have enough gifted students in a grade to fill an entire classroom. [6] Instead, most offer

[6] Typical class size for fourth and fifth grade is about 22-24 students, and is similar for gifted and non-gifted classes. In school year 2007-08 the average number of gifted fourth grade students at the 140 regular elementary schools in the District was 8.5. Only 8% of the schools had over 20 gifted students in fourth grade.

"mixed" classes that include gifted students and high-achievers. For simplicity we refer to both 100% gifted classrooms *and* mixed classrooms as "gifted classrooms."

Teachers of gifted classes must complete a 5-course sequence for state certification in gifted education. They are also slightly more experienced than their colleagues in regular classrooms. Students in a typical gifted classroom are highly selected, with standardized test scores about 0.75 standard deviation units higher than those of students in other classes at the same school. Nevertheless, all students in the District cover the same curriculum and write the same statewide achievement tests each year. Interviews with gifted teachers suggest that most of them divide their classes into ability groups (combining gifted and high achieving students with similar abilities), and assign enrichment projects for students who complete the regular curriculum material more quickly.[7]

Until 2004, potential candidates for gifted status were identified through parent and teacher referrals. Recommended students were referred to a District psychologist for a free IQ test. Parents could also pay for testing by a private psychologist and submit the test results to the school. Students with IQ's above the relevant threshold are eligible for gifted status, with final determination made in consultation between parents, teachers, and the school's Exceptional Student Education (ESE) specialist. The ESE specialist also draws up an Individualized Education Plan for each gifted student that specifies learning goals and instructional plans tailored to the student's strengths and weaknesses.

The IQ cutoffs used by the District for Plan A and Plan B students are dictated by state law. The law also allows students whose scores are within a standard error of the Plan A threshold (roughly 3 points) to be classified as gifted in cases where there is "overwhelming evidence" of superior ability.[8] As we discuss below, this is most likely to happen for students who score 129 points. There is no similar

[7] Our interviews suggest that few teachers cover advanced material from the next grade (so-called "acceleration"). See VanTassel-Baska and Brown (2007) for a discussion of general principles of gifted programming.

[8] A state interpretative memo states: " Eligibility committees may also judge a student eligible if there is overwhelming evidence in favor of more liberal interpretation, that is, by considering the standard error of measurement." See Florida Department of Education (1996).

allowance in the law for Plan B students, though we see a few disadvantaged students with scores of 115 points who are classified as gifted.

In response to concerns about the low numbers of disadvantaged students in gifted education, the District introduced a universal screening program in 2005. Under the program all second graders completed the Naglieri Non-verbal Ability Test (NNAT). Disadvantaged (FRL/ELL) students with a score of at least 115 points and non-disadvantaged students with a score of at least 130 points were referred to a District psychologist for IQ testing.[9] As in earlier years teachers and parents could still recommend students for testing, and parents could submit scores from private testing agents. Comparisons across cohorts of third-graders shown in Figure 1 suggest that the screening program raised the gifted fraction of disadvantaged students from under 0.5% to over 2%. A financial crisis in 2007 caused the District to cut funding for IQ testing, leading to a sharp drop in the placement rate of Plan B students. In 2010 the screening program was suspended altogether, resulting in a further fall in Plan B participation. A new screening test was introduced in 2011 but so far there is no indication of a return to earlier placement rates. Given the wide variation over time in Plan B enrollment rates, in our analysis below we test for potential differences between Plan B participants in the first three years of our sample (who entered fourth grade from 2005 to 2008) and those in later cohorts.

Figure 2 shows how the composition of gifted fourth grade classrooms varies across schools in the District with higher and lower fractions of FRL students.[10] We distinguish four groups of gifted classroom participants: Plan A and Plan B gifted students, advantaged (non-FRL/ELL) high achievers, and disadvantaged (FRL/ELL) high achievers. As expected, Plan A gifted children are concentrated at schools with lower enrollments in free and reduced price lunch programs, whereas Plan B children are

[9] For screening purposes the District used a rescaled version of the NNAT with mean 100 and standard deviation 15. The NNAT is designed to measure IQ but is known to be relatively noisy -- see Lohman et al. (2008).

[10] The 140 schools included in our analysis have at least 50 students on average in third grade (in years when the school is open) and are not classified as charter schools. The FRL fraction is strongly positively correlated (ρ=0.8) with the fraction of black non-Hispanic students in a school and strongly negatively correlated (ρ=-0.9) with average third grade scores in reading and mathematics.

concentrated at schools with moderate to high levels of FRL participation. Even at schools with FRL participation below 20%, however, only about one-half of the students in a typical gifted classroom are gifted. The majority of seats in the gifted classrooms in the District are filled by high achievers.

III. Student Achievement Data and Analysis Samples

We use administrative data for students who were in first through third grades during the period from 2004 to 2011 to study the impact of gifted education. The data include gender, race, ethnicity, FRL/ELL status, NNAT screening test scores (for second-graders from 2005 to 2009), and IQ test scores (for students who were tested). In addition, we have access to statewide achievement tests in reading and math that are mandated for all third and higher grade students, as well as writing tests for fourth and eighth grades and science tests for fifth grade. We use the fourth grade test results as our main outcome measures, and third grade scores as a measure of baseline (i.e., **pre**-program) achievement for students who participate in gifted classrooms.

A potential concern with the use of third grade scores as a baseline is that some gifted students have received special services (e.g., individualized instruction) in second or third grade. For many students in our sample we also have access to Stanford Achievement Tests (SAT's) in reading and math for first through third grades.[11] As a robustness check, we use these tests to measure baseline achievement *before* gifted students have been exposed to any form of gifted programming. As discussed below, we find very similar results using either set of scores, and in the interests of maximizing our sample sizes we use the third grade scores as the baseline measures for our main specifications.

Table 1 shows the overall characteristics of students who appear at least once in first through third grades at the 140 regular elementary schools in the District, as well as the characteristics of the samples we use to study the impacts of gifted education on the three groups of participants in gifted

[11] The District required SAT math tests in every grade until 2007 and the SAT reading tests until 2008.

classrooms.[12] Data for the overall student body in column 1 show that the District is highly diverse, with 30% white non-Hispanics, 36% black non-Hispanics, 27% Hispanics, and 4% Asians. One-half of students are eligible for free or reduced price lunches, 12% are English language learners, and 55% are either FRL or ELL and therefore fall under the Plan B rules for gifted eligibility. We use 2000 Census data on median family incomes by zip code to proxy family income: the average level is $57,800. Overall, about 6% of students in the available cohorts were classified as gifted by fourth grade and 13% were assigned to a gifted classroom.

The lower rows of the table report mean IQ and NNAT scores, and mean scores on third and fourth grade statewide tests. We also show the mean third grade scores of students' fourth grade classmates. IQ and NNAT test scores are missing for many students, so we report (in italics) the percentages of students with valid scores on each test. Both tests are scaled to have mean 100 and standard deviation 15 in a national population. We standardize the statewide achievement test scores to have mean 0 and standard deviation 1 within a grade/year cohort in the District.[13]

Columns 2 and 3 of Table 1 show the characteristics of our main analysis sample for studying Plan A participants in gifted education, distinguishing between students who were placed in a gifted classroom in fourth grade (column 2) and those who are not (column 3). The sample includes students who are neither FRL nor ELL and have an IQ test by the time they are entering fourth grade within 10 points of the 130 point IQ cutoff for Plan A eligibility. To sidestep problems of selective retesting we use each student's first recorded IQ test to define our sample. We also limit attention to students

[12] We restrict the cohorts to students who would potentially appear in fourth grade between 2005 and 2012. As discussed below, we find no evidence that gifted classroom assignment affects attrition from the Distrct.

[13] Since scores at the regular elementary schools are higher than scores at the charter and special schools in the District, the mean standardized scores for the sample in column 1 are slightly positive.

whose first recorded test was administered by a District psychologist to avoid selective reporting of IQ scores from private testing agencies.[14]

Comparisons between columns 2 and 3 of Table 1 show that the demographic characteristics of students with IQ scores around 130 points who were placed or not placed in a gifted classroom are quite similar. Both groups are more likely to be white and non-Hispanic than the District population as a whole, and are also drawn from relatively rich zip codes. As expected students who are placed in a gifted classroom (column 2) have higher IQ scores, NNAT scores, and third grade test scores than those who are not (column 3). The gap in statewide achievement tests between the placed and non-placed tends to *narrow* slightly between third and forth grades, however, showing no indication that placement in a gifted class leads to achievement gains. (For example, the gap in math scores between placed and non-placed students narrows from 1.45-1.09=0.36 to 1.39-1.10=0.29).

The bottom rows of Table 1 show the mean third grade math and reading scores of each groups' fourth grade classmates. As shown in column 3, even the students in our Plan A sample who are placed in regular classes have relatively strong peers (with mean third grade scores around 0.30), reflecting their concentration at higher-income schools with above-average test scores. Those in gifted classrooms have *very strong* peers, with mean third grade scores around 1.15 standard deviation units above the average for the District as a whole.

Columns 4 and 5 summarize our analysis sample for studying Plan B participants. These students are either English learners or eligible for free/reduced price lunch, and have a first-reported IQ score between 105 and 125 points. Relative to the overall District population they are less likely to be white non-Hispanic and more likely to be Hispanic. They are also drawn from somewhat below-average income zip codes. Their IQ's and third grade scores are lower than those of the Plan A analysis sample but still substantially above the district-wide averages.

[14] Parents have little or no incentive to report private tests below 130 points and nearly all private test scores recorded by the District are above 130.

As with the Plan A sample, students in the Plan B sample who are placed in a gifted classroom in fourth grade have higher third grade scores than those who were not. By the end of fourth grade the gaps between placed and unplaced students have narrowed slightly, again providing no indication that placement in a gifted class leads to achievement gains.[15] The peers of the Plan B sample members who are not placed in a gifted class have slightly below-average third grade scores, reflecting the fact that Plan B students tend to come from disadvantaged schools, while the peers of those who are placed in a gifted class have relatively high scores, though not as high as the Plan B students themselves.

Finally, columns 6 and 7 show the characteristics of our sample for studying high-achievers. To construct this sample we first identified schools with a fourth grade gifted classroom containing at least one non-gifted student. The District specifies that schools should assign open seats in a gifted classroom to the non-gifted students with the highest scores on the previous year's statewide tests (e.g., using third grade scores to determine fourth grade status). Prior to 2009, schools had some discretion over the ranking formula, but starting in 2009 the District imposed a uniform rule. We identified a subset of schools that appear to have closely followed the District rule for fourth graders in 2009-2012. We then calculated school-specific cutoff scores for admission to the fourth grade gifted class and selected the first 10 students with scores above this cutoff and the first 10 with scores below it.[16]

Comparisons with the means in column 1 show that high achievers are closer to the District wide population in terms of free lunch status and neighborhood income than either gifted group. Their third grade test scores are lower than those of the Plan A sample, but close to those of the Plan B students – a little higher in reading and a little lower in math. Their rescaled NNAT scores, on the other hand, are roughly 10 points (2/3 of a standard deviation) below those of the Plan B sample.

---

[15] For example, the gap in third grade math scores is 0.56, while the gap in 4th grade scores is 0.51.

[16] The cutoff for each school in each year is determined by the score of the N^{th} ranked non-gifted student, where N is the number of non-gifted students in the gifted classroom. As explained in the Data Appendix, in cases where missing 3rd grade scores lead to ambiguity in the cutoff, we choose a cutoff score to minimize the misclassification rate of students whose scores are outside an interval around the potential threshold. Note that in some schools there may not be 10 students above the cutoff (if fewer than 10 seats were available in the gifted class).

The gaps in baseline scores between high achievers who enter a gifted classroom and those who do not are smaller than for the Plan B sample (0.31 in third grade reading and 0.14 in math). After one year of participation the gap in reading is the same but the gap in math is wider (0.26), suggesting that the program may have raised math achievement. The peers of the high achievers in regular classes have about-average third grade scores, whereas those who are placed in gifted classes have relatively strong peers, with scores about 1 standard deviation above average for the district as a whole.

IV. Validity of RD Design and First-Stage Relationship

a. Evidence of Manipulation of IQ Scores

The validity of an RD design may be compromised by manipulation of the running variable (see e.g., Lee, 2008; McCrary, 2008; Lee and Lemieux, 2010). In the case of IQ tests, psychologists have some discretion in assigning scores and can potentially boost marginal students above the gifted threshold.[17] Figures 3a and 3b show the frequency distributions of scores for our Plan A and Plan B evaluation samples.[18] Despite the fact that we use only first-reported IQ scores from District-administered tests, both histograms show evidence of manipulation, with spikes at the minimum threshold scores for each group and deficits below the thresholds.

To assess the magnitude of the manipulation of the Plan A scores we fit a simple model to the histogram in Figure 3a, assuming that the frequency distribution of true scores is quadratic between 120 and 140 points, and allowing arbitrary fractions of students who score between 127 and 129 points to be "bumped up" to 130 or 131 points. The fitted model, shown by the lighter bars in the figure, provides a reasonable approximation to the true distribution, apart from scores of 125 and 126, which

[17] For example, on the "similarities" section of the Wechsler Intelligence Scale for Children (4th edition), children are given two words – like "red and blue" – and are scored 0, 1, or 2 points based on a qualitative assessment of their explanation for their similarity.
[18] The distribution of ranks in our high-achiever sample is uniform (by construction) so we do not show it.

are over-predicted by the model.[19] We also superimpose the implied frequency distribution from the fitted model, assuming no manipulation. Relative to this counterfactual, the model implies that about 65% of students with true scores of 127 or 128, and 82% of those with a true score of 129, are bumped up to 130 or 131 points. Consequently, 64% of the observations at 130 points and 36% at 131 points are attributable to students with true scores of 127-129 points.

We fit a similar model to the Plan B histogram in Figure 3b, assuming that arbitrary fractions of test takers who score 114 or 115 are bumped up to 116 or 117 points. As shown by the fitted histogram (plotted with lighter bars) the model yields a plausible fit, though a formal test rejects the model at conventional significance levels (chi-square=36 with 15 degrees of freedom). We also superimpose the implied frequency distribution in the absence of manipulation. Relative to this benchmark 18% of students who score 114 points and 52% of those who score 115 points are bumped up to 116. (There is little estimated displacement to 117 points). The model implies that 38% of students with a 116 score have true scores of 114 or 115.

While there is clearly some manipulation of IQ scores, the key concern for an RD analysis is whether this leads to systematic differences in the latent abilities of students on either side of the threshold. If IQ is imperfectly measured and students in a narrow range of scores are very similar, then *randomly* raising the scores of some students from just below the threshold to just above will not necessarily violate the conditions for a valid RD design. Non-random manipulation can be tested by looking for discontinuities in pre-determined characteristics like previous test scores. Figure 4 shows the relationship between IQ and four key background variables: median household income, the NNAT screening test, and third grade reading and math scores. Looking first at the Plan A students, none of the variables exhibits a large discontinuity at the 130 point threshold, although there is a small upward

[19] The goodness of fit statistic is 82.2 with 14 degrees of freedom, which is highly significant. Interestingly, however, we cannot reject the hypothesis that the number of under-predicted observations in the 127-129 score range equals the number over-predicted in the 130-131 score range. If we extend the range of scores that are bumped up to 125 points we find too many scores missing from below the threshold relative to the spikes above.

jump in math scores. (Formal test statistics for the baseline reading and math scores are discussed in the next section). Further evidence is presented in the first two columns of Appendix Table 1, where we show the difference in means of various characteristics for students just above and just below the Plan A threshold (i.e., with IQ's of 130-131 points versus 127-129 points), as well as coefficients from a logit model for the probability of being above the threshold, conditional on being in the 127-131 point range. None of the differences in means of the individual characteristics is significant at the 5% level, though third grade math scores are marginally significantly higher for students above the 130 point threshold (t=1.8), as are SAT math scores measured prior to the students receiving any gifted services (t=1.6). Likewise, lagged test scores and other characteristics are not jointly significant predictors of being above the gifted threshold (p-value=0.38). These comparisons suggest that the assignment of students to above or below the gifted threshold may be almost "as good as random," with any bias likely to lead to an *overstatement* of the positive effect of gifted participation. As a result we present conventional RD results for Plan A students below, including models that control for previous test scores. We also present donut-hole specifications that ignore students in the 127-131 point range, and first differenced models that are potentially more robust to manipulation.

Looking at the plots in Figure 4 for the Plan B analysis sample, again there are no obvious discontinuities in the four background characteristics. The similarity of students on either side of the 116 point threshold is confirmed in Appendix Table 1 (columns 3-4) where we compare mean characteristics for those with IQ's of 116 and 117 points versus those with IQ's of 114 or 115, and present logit coefficients from a model predicting who is above the threshold. As in the Plan A sample, none of the mean differences is significant, nor are the characteristics jointly significant in a logit model for scoring above the threshold, suggesting that the manipulation of scores is close to random. Again in our analysis below we supplement conventional RD models for the Plan B students with donut-hole specifications and first differenced models.

Although we have no reason to suspect any manipulation in the scores that are used to rank non-gifted students and determine eligibility for the gifted class, we also checked that the characteristics of the high achievers evolve smoothly as their rank passes the eligibility threshold. The patterns in Figure 4 confirm that median household income, NNAT scores, and third grade reading and math scores are all quite smooth at the cutoff rank for placement in a gifted classroom. These visual impressions are confirmed by the test statistics presented below.

b. Differential Attrition

To be included in our Plan A and Plan B analysis samples a student has to receive an IQ test from a District psychologist in first through third grades, remain in a District school through the end of fourth grade, and be promoted one grade per year. Differential attrition could arise if students who achieve gifted status are more likely to remain in the District, posing a threat to our RD analysis. A similar bias could affect our analysis of high achievers if students decide whether to leave the District after learning whether they are placed in the fourth grade gifted classroom.

We conducted a series of RD analyses, summarized in Appendix Figure 1, to check these concerns. For Plan A and Plan B students we identified all students with a first-time District-administered IQ test in first through third grades from 2002 to 2011, and examined the relationship between IQ scores and the probability of being included in our fourth grade sample. For high-achievers (who learn their placement status after enrolling in fourth grade) we examined the relationship between the rank at the start of fourth grade and the probability of remaining in the District through the end of the school year. We find no evidence of discontinuities in retention rates for any of the three groups. Based on these results and a series of RD style models (available on request) we conclude that differential attrition is not a major concern for our analysis.

c. First Stage Relationships for Gifted Status and Placement in a Gifted Classroom

With this background we turn to the first-stage relationships between IQ scores (or test score ranks) and participation in gifted programs. For Plan A and Plan B students we consider two measures of participation: achieving gifted status by fourth grade and placement in a gifted classroom in fourth grade. For our high achiever analysis sample we consider only placement in a gifted classroom.

Figure 5a shows the first-stage relationships for our Plan A sample. We plot the average fraction of students classified as gifted and the average fraction placed in a gifted classroom at each test score value, and the fitted relationships from a pair of fully interacted linear models. To the left of the threshold the fraction placed in a gifted classroom is higher than the fraction gifted, reflecting the fact that many high-IQ students are also high achievers and enter a gifted class through that route. Both series are also rising with IQ. Some of the rise is caused by retesting: students with a higher initial IQ score are more likely to be retested (often privately) and end up classified as gifted. Some is also due to the provision in state law that allows students with scores below 130 to be eligible for gifted status. This is most likely to happen for students with a score of 129 points. As noted in Figure 3a, however, there are very few Plan A students with a score just 1 point below the threshold and their presence has little impact on the local linear models. To the right of the 130 point threshold the fractions of students classified as gifted and participating in a gifted classroom are both very close to 100%. The local linear models imply a first stage discontinuity in the probability of being classified as gifted of about 50 percentage points, and a discontinuity in the probability of participating in a gifted class of about 25 percentage points.

Figure 5b shows similar first stage models for Plan B students. To the left of the 116 point threshold the fraction classified as gifted is low and exhibits little upward trend, suggesting that very few FRL/ELL students who miss the gifted threshold are re-tested. The fraction of students with scores from 105 to 115 who are placed in gifted classes, however, is increasing in the IQ score, reflecting the rising likelihood of being placed in a gifted classroom as a high achiever. There is also a rise in the placement rate in a gifted class for students who score 1 point below the threshold (i.e, at 115 points). The local linear models show clear discontinuities at the eligibility threshold in the probability of being classified as a gifted student (about 50 percentage points) and in entering a gifted classroom in fourth grade (roughly 35 percentage points).

While nearly all Plan A students who score 130 or more points on an IQ test are classified as gifted (and placed in a gifted classroom), only 55% of marginally eligible Plan B students achieve gifted status. As we discuss below, the gifted placement rate for marginally eligible Plan B students is particularly low at richer schools with few FRL students. Some of this may be due to teachers' concerns that marginally eligible Plan B students will be "mismatched" to the level of the other students in the gifted classrooms at these schools. In the models below we therefore consider interacted models that allow different effects on Plan B students at higher and lower FRL schools.

For high achievers, Figure 5c shows that the first stage relationship between their relative rank in the pool of high-scoring non-gifted children and their probability of participation in a gifted classroom is similar to the corresponding first stage relationship for Plan B's, with a clear discontinuity at the cutoff score (about 35 percentage points). The "non-compliance" behavior for high achievers (i.e., positive participation rates for students below the cutoff and less than 100% rates for students above the cutoff) is attributable to several factors, including missing test score data, movements of students in and out of schools, and non-compliance with the District formula. Unlike the Plan B students, however, compliance rates do not seem to vary systematically across schools with higher and lower FRL rates.

d. Discontinuities in Peer Quality

An important channel through which assignment to a gifted classroom could affect student learning is through peer effects, which some analysts believe are important inputs in the education production process (see Sacerdote 2011 and Angrist 2013, for recent reviews). Figure 6 shows the relationships in our three analysis samples between IQ or test score rank and peer quality, as measured by mean third grade reading and math test scores of classroom peers (panels a and b), the mean fraction of gifted peers (panel c), and the mean fraction of peers with a learning disability (panel d). For all three analysis samples we find positive discontinuities in the mean lagged test scores of classroom peers (on the order of 0.2 standard deviation units), positive discontinuities in the fraction of peers who are gifted, and negative discontinuities in the fraction of peers with a learning disability. For Plan A students the rise in the fraction of gifted peers around the 130 point IQ threshold is relatively large and closely tracks the probability of being placed in a gifted classroom (Figure 5a). For Plan B students and high achievers the discontinuity in the fraction of gifted peers is smaller, since most of these students are in schools where the majority of students in gifted classrooms are high achievers. Nevertheless, all three measures suggest a gain in peer quality for students who are placed in a gifted classroom.

V. Impacts on Student Achievement

a. Framework

The District's gifted program causes variation in three features that may affect test scores: placement in a gifted classroom (with a certified teacher, a faster pace of instruction, and more time for enrichment projects), receipt of individualized gifted services, and quality of classroom peers. A very simple structural model that incorporates all three factors is:

$$(1) \qquad y = \beta_1 D_{class} + \beta_2 D_{gifted} + \beta_3 Q_{peer} + \beta_x X + \lambda\, IQ^* + \varepsilon$$

where y represents the test score (in some domain) for a given student at the end of fourth grade, D_{class} is an indicator for being placed in a gifted class, D_{gifted} is an indicator for being classified as gifted, Q_{peer} is a measure of the relative quality of a student's classroom peers, X represents a set of observed covariates (e.g., gender, race/ethnicity, school dummies), IQ^* is the student's true cognitive ability (as would be measured by a non-manipulated IQ test), and ε is an error term reflecting variation in test performance and other factors. Taking expectations conditional on observed IQ:

(2) $$E[y|IQ] = \beta_1 P[D_{class}=1|IQ] + \beta_2 P[D_{gifted}=1|IQ] + \beta_3 E[Q_{peer}|IQ] + E[\beta_x X + \lambda IQ^* + \varepsilon|IQ]$$

Assuming that the X's vary smoothly at the gifted threshold (as in Figure 4), and that $E[IQ^*|IQ]$ is continuous at the threshold (as would be true if the manipulation of observed IQ scores is ignorable), the discontinuity in the conditional expectation function $E[y|IQ]$ at the gifted threshold (T) is:

(3) $$\mathbf{Dis}(y) = \lim_{IQ\downarrow T} E[y|IQ] - \lim_{IQ\uparrow T} E[y|IQ] = \beta_1 \mathbf{Dis}(P_{class}) + \beta_2 \mathbf{Dis}(P_{gifted}) + \beta_3 \mathbf{Dis}(Q_{peer}),$$

where $\mathbf{Dis}(P_{class})$ and $\mathbf{Dis}(P_{gifted})$ are the discontinuities in the probabilities of being placed in a gifted class and classified as gifted, respectively, and $\mathbf{Dis}(Q_{peer})$ is the discontinuity in peer quality. A similar model is relevant for high achievers using test score rank as the running variable and setting $D_{gifted} = P_{gifted} = 0$.

In our empirical analysis we focus on the reduced form discontinuities in test scores at the eligibility threshold, which we interpret as estimates of $\mathbf{Dis}(y)$. For some purposes it is useful to scale the reduced form effect by dividing by the estimated discontinuity in the probability of being placed in a gifted class, providing estimates of the ratio:

$$\mathbf{Dis}(y) / \mathbf{Dis}(P_{class}) = \beta_1 + \beta_2 [\mathbf{Dis}(P_{gifted}) / \mathbf{Dis}(P_{class})] + \beta_3 \mathbf{Dis}(Q_{peer}) / \mathbf{Dis}(P_{class}) .$$

The left hand side is the standard "treatment on the treated" effect assuming that treatment operates through assignment to a gifted class. When $\beta_2 = \beta_3 = 0$, this is equal to the causal effect of being placed in a gifted class, β_1. When $\beta_2 = 0$ but $\beta_3 \neq 0$, the rescaled reduced form effect also includes a peer effect component, consisting of the treatment on the treated effect on peer quality for those who move to a gifted class ($\mathbf{Dis}(Q_{peer})/\mathbf{Dis}(P_{class})$), multiplied by the causal effect of peer quality β_3. Importantly, we can compare reduced form impacts in different subsamples where we observe different values for $\mathbf{Dis}(Q_{peer})/\mathbf{Dis}(P_{class})$, allowing us to make inferences about the likely importance of the peer quality component in the combined treatment effect.

When there is also a causal effect of being classified as gifted (i.e., $\beta_2 \neq 0$) the rescaled reduced form effect includes a third component attributable to this effect. We suspect that β_2 is small, however, since the individualized services offered to gifted students are only for a few hours per week, and are not focused on mastering the standard curriculum. Indeed, comparisons between the third grade statewide test scores of gifted students and their scores on SAT tests written prior to the receipt of any gifted services show no evidence that third grade scores are affected by individualized instruction offered prior to fourth grade.

Equation (2) also makes clear the potential threat to an RD analysis if there is non-random manipulation of measured IQ scores. If, for example, the true cognitive ability of students whose IQ scores are boosted to just above the gifted threshold is greater than the true ability of students whose scores are not boosted, $E[IQ^{*}|IQ]$ will exhibit a positive discontinuity at IQ=T, implying an upward bias in the reduced form discontinuity. As we discuss below, if the additive specification of equation (1) is correct -- as appears to be true in our data -- this bias can be eliminated by first-differencing.

b. Plan A Students

The reduced-form relationships between IQ and fourth grade test scores in reading, math, and writing for our Plan A analysis sample are shown in Figures 7a-c. As in Figures 4-6, we show the mean outcomes for each single IQ point, and the fitted relationship from linear regression models fit to the two sides of the threshold separately. The graphs suggest an upward-sloping relationship between IQ and average test scores, with small negative discontinuities at the gifted threshold for reading and math a small positive effect for writing.

Table 2 presents the estimated discontinuities at the 130 point threshold from a variety of alternative RD specifications. (We defer for the moment the issue of bandwidth selection, and focus on results from models that use 10 points to the left of the IQ threshold and 11 points to the right). Each column corresponds to a different dependent variable, whereas each row corresponds to a different specification of the RD model, including models with no added controls (row 1), models with a broad set of controls including school dummies (row 2), donut-hole specifications that exclude data within 2 points of the threshold (row 3), and first differenced models (rows 4 and 5). Columns 1 and 2 take as dependent variables the baseline third grade test scores in reading and math.[20] The next three columns present first stage models for the event of being placed in a gifted classroom in fourth grade (col. 3), being classified as gifted by fourth grade (col. 4), and the average baseline test scores of the classroom peers in fourth grade (col. 5).[21] Finally, columns 6, 7, and 8 present reduced-form models for standardized scores in fourth-grade reading, math, and writing tests, respectively.[22]

The estimated discontinuities shown in columns 1-2 of Table 2 suggest that the relationship between the baseline test scores and IQ is relatively smooth at the 130 point threshold, with a very small jump in reading scores (0.02 to 0.03 standard deviation units) but a slightly larger jump in math scores (around 0.09 standard deviation units). The higher math scores for those above the threshold are consistent with the simple comparisons in Appendix Table 1, and suggest that there may be some positive selection bias in simple RD models for math achievement.

The estimated first stage discontinuities in the probabilities of placement in a gifted class, being classified as gifted, and in peer quality (columns 3-5) are all insensitive to the addition of controls for school dummies and student characteristics (compare rows 1 and 2), but become slightly larger when data close to the threshold are eliminated (row 3), as expected given the patterns in Figure 5.

With the exception of the model in row 1 for writing scores, all the reduced form discontinuity estimates in columns 6-8 are negative. The point estimates become more negative when student characteristics and school fixed effects are added (row 2), and are even more negative in the donut-hole

[20] The models in row 1 correspond to the fitted models for Plan A students shown in panels C and D of Figure 4.
[21] The models in row 1 for being classified as gifted or placed in a gifted classroom correspond to the fitted models shown in Figure 5a.
[22] The models in row 1 for these outcomes correspond to the fitted models shown in Figure 7a-c.

specifications, but also more imprecise, reflecting the reduction in sample size and the limited range of the RD running variable in these models.

The first differenced (or "gain-score") models in row 4 of Table 2 use as a dependent variable the change in standardized scores in reading and math from third to fourth grade. (Since writing is not tested in third grade we cannot estimate differenced models for this outcome). Assuming that equation (1) is correctly specified for both third and fourth grades and that $D_{class} = D_{gifted} = 0$ for all students in third grade, we can difference the model, obtaining:

(4) $$Dy = y - y_0 = \beta_1 D_{class} + \beta_2 D_{gifted} + \beta_3(Q_{peer} - Q_{peer3}) + \varepsilon - \varepsilon_0 ,$$

where y_0 represents third grade achievement, Q_{peer3} is peer quality in third grade, and ε_0 is the value of the unobserved error component in the model for third grade achievement.[23] Assuming that peer quality in third grade varies smoothly with IQ, the discontinuity in the conditional expectation of Dy given IQ at the threshold for gifted eligibility is:

(5) $$\mathbf{Dis}(Dy) = \lim_{IQ\downarrow T} E[Dy|\ IQ] - \lim_{IQ\uparrow T} E[Dy|\ IQ] = \beta_1\, \mathbf{Dis}(P_{class}) + \beta_2\, \mathbf{Dis}(P_{gifted}) + \beta_3\, \mathbf{Dis}(Q_{peer}),$$

which is the same as the discontinuity in the level of test scores under the assumption that the basic RD model is unbiased. Thus, a comparison of results from our basic RD model and the alternative first differenced model provides a test for the effect of non-random sorting around the gifted threshold.

The key feature of the differenced specification in equation (4) is that the gain in test scores is unrelated to IQ. Under this assumption, even if the manipulation of IQ scores is non-random so a conventional RD model is biased, the differenced specification provides a valid estimate of the reduced form impact of interest. Importantly, the maintained assumption of this model can be tested by checking that the growth in test scores does not depend on initial IQ (except possibly at the gifted threshold).

---

[23] For simplicity we ignore any time-varying effect of the student covariates in equation (4), but these are all included in our estimating models, as are school fixed effects. We have also ignored the fact that some gifted students receive individualized instruction and other gifted services in third grade. Assuming that these services exert a constant effect on achievement, differencing across grades will difference out their effect, eliminating the $\beta_2 D_{gifted}$ term for students who were identified as gifted prior to third grade.

Figures 7d and 7e show the relations between the gains in reading and math scores and IQ in our Plan A sample, along with the fitted values from linear models fit separately to each side of the 130 point threshold. To provide additional leverage we expand the samples in these figures to include students with IQ scores from 115 to 145 (i.e., a bandwidth of 15 points). Consistent with equation (4), the figures show that initial IQ has very little relationship with the change in reading or math scores. Moreover, there is no evidence that students with IQ scores above 130 points tend to experience larger gains than students with scores below 130. These impressions are confirmed by the estimated models in row 4 of Table 2, which suggest impacts of placement in a gifted class that are small and insignificant. For completeness in row 5 we present differenced models that exclude students with IQ scores from 127 to 131 points. The estimates from these models are less precise but again show no evidence of faster achievement growth for students with IQ scores above 130.

Overall, the estimates in Table 2 suggest that participation in gifted education has little or no effect on standardized test scores of Plan A students. In fact the point estimates from most of the specifications are negative, though we typically cannot reject small positive effects.[24] We have conducted a wide range of robustness checks on these results, including RD models with alternative bandwidths, and models that use SAT scores written prior to receiving any gifted services to measure baseline achievement. The results from models with alternative bandwidth choices are summarized in Appendix Figure 2, where we show point estimates and standard error bands for the estimated impacts on reading, math and writing using the simple specification in row 1 of Table 2 with bandwidths from 5 to 15 points. For reading, the point estimates become more negative as the bandwidth is increased, with a marginally significant -0.1 estimate using bandwidths of 13 or larger. For math and writing the point estimates tend toward a value closer to 0 as the bandwidth is increased, but remain uniformly negative.

[24] For example, taking our most precise models (in row 2) we cannot reject an effect size of 0.05 on reading, 0.04 on mathematics, or 0.08 on writing using conventional significance levels.

Results from models that use SAT scores measured before a student is classified as gifted as a baseline measure of achievement are presented in Appendix Table 2. These results are quite similar to the results in Table 2, though typically less precise, reflecting the smaller sample sizes. We have also examined effects on fifth grade student outcomes, capturing the effects of two years of exposure to a gifted classroom. These effects (reported in Appendix Table 3) are also small in magnitude and mostly. Finally, we have examined RD models for various subsamples of students, based on school characteristics such as the fraction of FRL participants or average test scores in the school as a whole. Given our limited sample sizes the results are somewhat imprecise, but we find no evidence of positive effects for any subsample.

An important feature of RD program estimates is that they identify causal impacts for *marginally eligible* subjects -- in our case, students around the threshold for placement in a gifted class. Bui et al. (2012) have conjectured that marginally eligible gifted students may be harmed by being placed in a gifted program where higher teacher expectations and higher-ability peers may lead to reduced self-esteem and motivation.[25] Some insight into the potential difference in impacts for marginally eligible Plan A students versus other very-high ability students is provided by the differenced estimates, which are based on difference-in-differences for *all students* with IQ's of 130 or more, versus *all those* with IQ's of 129 and below. (See also panels D and E of Figure 7). The fact that these estimates are so similar to the RD estimates suggests that even students with IQ's well above 130 show little or no gain in achievement relative to the trend in growth for students with IQ's of 120-129.

Additional evidence on potential mismatch of marginally eligible Plan A's is provided by analyzing responses to annual surveys conducted by the District to assess students' satisfaction with their learning environment. (See the Data Appendix for more information on this survey). We examine

[25] In most gifted classrooms even marginally eligible Plan A students have higher ability than their peers (who are mainly high achievers). Students with IQ's in the range of 127-131 points have standardized average third grade scores of about 1.3 (see Figure 4), whereas on average their gifted class peers have scores of about 1.15 (Figure 6).

the relationship between IQ scores and the probability that a student "strongly agrees" with each of four statements: (1) "my teacher(s) believe I can succeed", (2) "my teacher(s) answer questions in a way I understand, (3) "I enjoy learning in my school" and (4) "I feel accepted/like I belong in my school."

Panel A of Appendix Figure 3 shows the relationship of survey responses to IQ scores, while columns 2-3 of Table 3 show results from fitting RD models to the responses to each of the four questions, and a composite average of all four responses. Although the estimates are imprecise, all four measures of student satisfaction exhibit *positive* discontinuities at the gifted eligibility threshold.[26] These patterns are the opposite of what we would expect if marginally eligible gifted students have difficulty understanding the material or experience invidious comparison effects. Instead, marginally eligible students appear to be *more satisfied* with their learning environment.

c. Plan B Students

Figure 8 shows the reduced form relationships between IQ and fourth grade test scores for our Plan B analysis sample. Mean reading and math scores evolve relatively smoothly through the gifted threshold at 116 points and provide little evidence of an effect of gifted services on these outcomes. Writing scores, however, exhibit a jump at the gifted threshold, suggesting a potentially positive impact.

Table 4 presents estimated discontinuities for various outcomes and model specifications, following the same format as Table 2. Notice first that the estimated discontinuities in baseline reading and math scores are small and insignificant, confirming the visual impression from panels C and D of Figure 4. The first stage discontinuities in the probabilities of being placed in a gifted class or being classified as gifted (columns 3-4) and in classroom peers' previous scores (column 5) are relatively precisely estimated and very similar whether we control for student characteristics and school dummies or not (compare rows 1 and 2). The donut-hole specifications (row 3) yield first stage estimates that are a little bigger in magnitude, consistent with the evidence in Figure 5b and 6 that the first stage

[26] Because not all students completed the survey (the response rate is around 95%) we also test for discontinuities in the response rate (row 6 of Table 3). The estimates are small in magnitude and statistically insignificant.

outcomes for students in the Plan B sample with IQ's just below the gifted threshold are higher than expected. Removing these observations therefore tends to widen the estimated discontinuities in the first stage outcomes.

As suggested by the patterns in Figures 8a and 8b, the reduced-form RD estimates for fourth grade reading and math in columns 6 and 7 of Table 4 are uniformly small in magnitude, regardless of whether we control for school effects and student characteristics or exclude observations within 2 points of the gifted threshold. In Appendix Figure 4 we explore the robustness of these basic RD results to alternative bandwidth choices. Across a range of bandwidths from 5 to 15 points, the estimated impacts on reading and math are uniformly close to zero.

Rows 4 and 5 of Table 4 show estimates from differenced models that use the change in test scores from third to fourth grade as the dependent variable. For additional leverage we expand the sample for these models to include students with IQ scores from 100 to 130 points. Figures 8D and 8E confirm that, as we saw for Plan A students, there is virtually no relationship between test score growth and IQ. Estimates from the differenced models confirm the conclusion from our basic RD models that there is no impact of placement in a gifted classroom on reading or math test scores of Plan B students.

In contrast, the results in column 8 of Table 4 show reduced form impacts on writing scores that are relatively large in magnitude (0.11 to 0.12 standard deviation units), stable across specifications, and at least marginally significant. As shown in Appendix Figure 4, the point estimates for writing impacts from the simple model in row 1 are also quite stable across alternative bandwidth choices. Our most precise specification in row 2 yields a t-statistic of 2.5 for the test of no effect of gifted services. Scaling the reduced form effect on writing scores by the first stage effect on being assigned to a gifted class (from column 3) implies a treatment effect per Plan B student who is assigned into a gifted classroom of about 0.3 standard deviation units—a relatively large effect.

To explore potential heterogeneity in the impacts of gifted services on Plan B students we fit a series of interacted RD models. These models take the form:

$$(5) \quad y = \beta_0 + \beta_x X + \gamma_0 1[IQ \geq T] + \gamma_1 (IQ - T) + \gamma_2 (IQ - T)\, 1[IQ \geq T] + \\ + \rho_0 1[IQ \geq T]\, Z + \rho_1 Z + \rho_2 (IQ - T)\, Z + \rho_3 (IQ - T)\, 1[IQ \geq T]\, Z + e ,$$

where 1[.] represents the indicator function, T is the gifted threshold (116 points for Plan B's), X is a set of controls, Z is a dummy taking the value of 1 for a specific subgroup of students and or schools, and e is an error term. Note that apart from the common coefficients for the X's these models are equivalent to estimating linear RD models separately for the Z=0 and Z=1 samples. We are interested in whether the implied discontinuity at the gifted threshold is different for the Z=1 and Z=0 subgroups, which can be tested by examining the interaction coefficient ρ_0.

Panel A of Table 5 presents estimation results from interacted models with four different "Z" variables: an indicator for early (2005-8 in fourth grade) versus late (2009-2012) cohort; an indicator for gender; an indicator for high or low FRL rate in the school; and an indicator for lower or higher peer quality of the gifted class. The cohort and gender interactions in the baseline achievement models, the first stage outcomes, and the outcome models for reading and math scores are all relatively small in magnitude and insignificant. The interactions of both cohort and gender with the discontinuity in writing scores, however, are larger in magnitude and marginally significant. Panels A-C of Figure 9 present graphical evidence on the differences in the relationship between IQ and test score outcomes by gender. The only discernable gender difference is for writing scores: boys appear to have relatively large gains at the gifted threshold, whereas the gains for girls are smaller.

Comparisons by cohort (not shown) are broadly similar. In particular, there is no evidence that that discontinuities in baseline scores, first stage outcomes, or fourth grade test scores in reading or math differ between earlier cohorts (who were in the relevant grades during the expansion of Plan B enrollment) and later cohorts (when fewer students were placed as Plan B gifted students). For writing scores, however, there is a strong interaction effect, with all of the discontinuity attributed to the early cohort. We also examined a gender × cohort interaction, focusing on early-cohort boys versus all other students. Here, as expected, we find that nearly all the discontinuity in writing scores arises from early cohort boys, with a reduced form effect of 0.27 standard deviation units (std. error=0.13).

The two other interaction models considered in Panel A of Table 5 have to be interpreted carefully because in both cases the first stage models vary with the interaction variable. As noted in the discussion of Figure 5b, Plan B students who score just above the 116 point threshold have higher probabilities of being placed in a gifted program and being classified as gifted at high-FRL schools. This pattern is confirmed by the relatively large and significant interaction effects in row 3 of Table 5 for the first stage models in columns 3 and 4. Nevertheless, the estimated interactions in reading and math outcomes are both relatively small, implying that there is little difference in the effect of gifted services on Plan B students' reading and math scores at richer and poorer schools. A similar pattern is present in the models that include interactions for higher and lower peer ability in the gifted class (row 4). At schools with below-average baseline scores in the gifted class there are larger discontinuities in the rate Plan B students are placed in a gifted class and classified as gifted, but there is little evidence of any differential effect on their reading or math scores.

In contrast, the interaction models for writing achievement (column 8) suggest that the discontinuity at the Plan B threshold is significantly larger at schools with a higher fraction of free- and reduced price lunch participants, and marginally bigger at schools with lower-quality peers in the gifted classroom. Notice that there is no interaction between the school-level FRL variable and the peer effect from entering a gifted classroom, yet there is a large interaction with the reduced form writing outcome. This suggests that the writing impact is not mediated through peer effects. On the other hand, the first-stage effect on peer quality is smaller in magnitude in schools where the mean peer quality in gifted classes is below-average (i.e. the interaction term is significantly negative) and yet the effect on writing scores is also bigger in these schools. Again we infer that the impact on writing cannot be attributed to a peer effect. Instead, the concentration of writing gains among early cohort boys (a group with relatively low writing achievement scores, as shown in Figure 9) suggests that the combination of gifted services benefited the lowest-achieving writers more than others.

In Panel B of Table 5 we address the question of whether the impacts at the end of fourth grade hold up over longer horizons. Specifically, we compare estimates for fourth and eighth grade test scores using the early cohort of Plan B students (in 4^{th} grade between 2005 and 2008) and allowing the effects to vary by gender. Graphical evidence on these comparisons is presented in Panels D and E of Figure 9. As with the fourth grade results, we see negligible impacts on eighth grade reading or math scores, coupled with a positive effect on writing achievement for boys. The 8th grade discontinuity for boys is 0.23 (adding the main effect and interaction) -- about 2/3 as large as the discontinuity in fourth grade scores. Overall, the longer run impacts are consistent with the pattern of effects for fourth grade.

Finally, as we did for the Plan A sample, we explore the effect of gifted participation on the self-reported satisfaction of Plan B students. The results are summarized graphically in Appendix Figure 3 (Panel B), with corresponding model estimates in columns 5 and 6 of Table 3. The patterns are strikingly different from the Plan A results and point to uniformly *negative* effects of gifted placement on all dimensions of student satisfaction. In particular, the estimates suggest that marginally eligible students are 5 percentage points (or about 10%) less likely to strongly agree that "my teacher answers questions in a way I understand" and 4-5 percentage points less likely to agree that they enjoy learning in their school or feel accepted at their school.

One potential explanation for these negative effects is that marginally eligible Plan B students are less well prepared academically than other students in their gifted classes -- a "mismatch" effect. Importantly, however, marginally eligible Plan B students **do not** have lower lagged achievement than their peers: their third grade test scores are actually slightly above the scores of their classmates in gifted classes. Plan B students who enter a gifted class **do** experience a drop in achievement relative to their peers which might trigger an invidious comparison effect. The drop for marginally eligible Plan B students (about -0.7 standard deviation units) is only slightly bigger than the drop experienced by Plan A students (who typically enter gifted classes with very high-achieving peers), and is *smaller* than the drop in relative achievement experienced by high achievers. Since Plan A students show a rise in satisfaction upon entering gifted classes, while high achievers show no effect one way or the other (see below), there does not seem to be a consistent link between relative achievement and student satisfaction.

d. High Achievers

Figure 10 plots the outcomes of students in our high achiever analysis sample by their relative rank on the previous year's tests within their cohort and school. The plots for fourth grade reading and math scores show clear jumps at the threshold for admission to the gifted classroom —a pattern that is mirrored in the plots for the first differences in scores. In contrast, the plot for fourth grade writing shows little evidence of a discontinuity at the threshold rank.

Table 6a shows the estimated discontinuities in our main outcome variables from alternative model specifications, using the same format as Tables 2 and 4. The results for the baseline test scores in columns 1 and 2 show only small changes in reading and math scores at the threshold rank. The first stage models for the high achievers in columns 3 and 4 of Table 6a are precisely estimated and quite robust to the addition of controls: these models show a relatively large (0.32) discontinuity in the fraction of high-achievers placed in the gifted class at the cutoff, and a slightly smaller discontinuity (0.28) in mean lagged achievement scores of their classroom peers.

The estimated reduced-form discontinuities in reading and math are also relatively precisely estimated and robust across alternative specifications, with a magnitude of 0.07 to 0.10 standard deviation units, while the effects on writing are close to zero. Appendix Figure 5 shows the robustness of these results to alternative bandwidth choices, using the simple specification in row 1 of Table 6a. Across a range of bandwidths from 5 to 15 points, the estimated impacts on reading and math are quite stable and statistically significant at conventional levels, while the estimated impacts on writing are uniformly small. Taking account of the first stage effect, the reduced form impacts on reading and math achievement imply treatment-on-the-treated effects of being placed in a gifted classroom of roughly 0.3 and 0.23-0.33 standard deviation units, respectively.

As with our Plan B sample, we estimated a series of interacted RD models that allow the first stage models and reduced form impacts to vary by gender, between schools with a higher or lower fraction of students in FRL programs, and between schools with higher- or lower mean prior achievement for students in the fourth grade gifted classroom. The results are summarized in Table 7. The gender interactions for reading are marginally significant and suggest a larger effect of placement in a gifted classroom for boys than girls. The estimated interactions for math also point to a larger effect for boys, though they are smaller than the corresponding estimates for reading, and not significantly different from zero at conventional levels. In contrast to the pattern observed for Plan B students, there is no indication of a larger effect on writing achievement for boys.

The interactions with an indicator for attending a high-FRL school show a potentially interesting pattern. The first stage models suggest that the discontinuities in being placed in a gifted class and in average peer quality are quite similar in low- and high-FRL schools. The estimated outcomes models, however, show some evidence of differential treatment effects at low- and high-FRL schools, with a larger impact on reading scores at high-FRL schools, and a larger effect on math scores at low-FRL schools. Again, as with gender, there is no evidence of heterogeneity in the effect on writing scores.

The estimation results for models that allow an interaction with an indicator for schools with a relatively low-quality peer group in the gifted class are also interesting. The first stage models show no interaction with the probability of placement in a gifted class, but (as expected) a negative and marginally significant interaction for average peer quality. Nevertheless, the models for reading scores show a *bigger* positive effect at these schools. This pattern would seem to rule out peer effects as the main driver of the overall positive effect on reading scores. The interaction effects on math scores are smaller in magnitude and differ in sign between the levels and value added models, suggesting no systematic variation in math effects by the quality of the gifted classroom peer effect. Again, we interpret these results as implying that the benefits of participating in a gifted classroom are not mediated through peer effects.

To what extent do the achievement gains of high achievers in fourth grade persist to later years? Since our analysis sample starts with the 2009 cohort of fourth graders, our ability to examine longer-term outcomes is limited, and we focus on fifth grade scores. A first concern is whether students are more (or less) likely to stay in the District until fifth grade if placed in a fourth grade gifted class, creating potential selection biases. As shown in column 1 of Table 6b there is no evidence of such an effect.

Columns 2 and 3 of Table 6b present first-stage models for the probabilities of placement in a gifted class in fourth and fifth grades, respectively, as a function of a student's rank on the third grade statewide tests. The first two panels of Appendix Figure 6 graph these first-stage relationships. Reassuringly, the estimates in column 2 are very similar to the corresponding estimates in column 3 of Table 6b, and show a jump of about 34-35 percentage points in the probability of placement in a gifted class in fourth grade at the cutoff score. The estimates in column 3 show a discontinuity in placement in a fifth grade gifted class that is about one-fifth as big. Since placement in fifth grade depends on *fourth grade* scores, this discontinuity is not mechanical, and is presumably driven in part by the achievement gains experienced by students who attended a gifted class in fourth grade. Given the discontinuities in both fourth and fifth grade participation, the estimated effects on fifth grade test scores represent a combination of an effect that persist from fourth grade, plus the effect of participation in fifth grade, with a relative coefficient on the latter effect of about one-fifth.[27]

Estimates of the reduced form models for fourth grade reading and math scores in columns 4 and 5 of Table 6b are very similar to the corresponding estimates for the overall analysis sample in columns 5 and 6 of Table 6a. These reduced form relationships are shown in panels c and d of Appendix Figure 6. The reduced form effects on *fifth grade* reading scores are relatively small and suggest that the achievement gains in reading registered at the end of fourth grade fade relatively quickly. (See panel e of Appendix Figure 6). In contrast, the estimated effects on fifth grade math scores are about the same size as the effects on fourth grade scores, and are either marginally significant or significant. The relatively large impact of fifth grade scores is visually evident in panel f of Appendix Figure 6. Perhaps most interesting is the effect on fifth grade science achievement, which is about the same size as the effect on fourth grade math, and (in the richer specification in row 2) significantly different from zero at conventional levels. Taken together the math and science results suggest significant persistence of at least the math achievement gains of high achievers.

Given the positive effects on reading and math achievement for high achievers who are assigned to a gifted classroom, it is informative to examine their responses to the student surveys. RD models for these responses are presented in columns 8-9 of Table 3: the underlying data are plotted in Appendix Figure 3, Panel C. The graphs and models suggest that, unlike their Plan A or Plan B counterparts, high achievers who are eligible for placement in the gifted classroom do not differ consistently from those who just miss the cutoff in reported satisfaction with their learning environment. Across the four

---

[27] Assume for simplicity that the discontinuity in fourth grade scores is $\mathbf{Dis}(y_4)=\beta_1\mathbf{Dis}(P_{class4})$ (i.e., all the effect arises though placement in a gifted class in fourth grade). A plausible model for the discontinuity in fifth grade scores is $\mathbf{Dis}(y_5)=\delta\beta_1\mathbf{Dis}(P_{class4})+ \beta'_1\mathbf{Dis}(P_{class5})$, where δ is the fraction of the effect in fourth grade scores that persists to the next year's scores, β'_1 is the direct effect of placement in a gifted class in fifth grade on fifth grade scores, and $\mathbf{Dis}(P_{class5})$ is the discontinuity in the probability in placement in a fifth grade gifted class at the cutoff for admission to the fourth grade class. Then the expected reduced form effect on fifth grade scores, as a fraction of the effect on fourth grade scores, is $\delta+\beta'_1\mathbf{Dis}(P_{class5})/ \beta_1\mathbf{Dis}(P_{class4}) \approx \delta+0.2\ \beta'_1/\beta_1$.

questions the estimated discontinuities are uniformly close to zero, and the index of all four responses is very smooth at the eligibility threshold. The absence of large effects on satisfaction levels of the high achievers is notable in light of the relatively large gains in peer quality experienced by those who enter a gifted classroom (a treatment effect on the treated of 0.8 standard deviation units). High achievers' satisfaction levels do not seem to be strongly affected by peer composition, or indeed by any other feature of the gifted classrooms in the District.

e. An Alternative Design for High Achievers

A potential concern with the policy of placing gifted and high-achieving students in a separate classroom is that moving these children can affect the scores of *other* children, leading to possible biases in our RD procedure. If, for example, removing the most able students has a positive effect on the highest-ability children who remain in the class, then our RD estimates could understate the causal effect of gifted placement on high ability children. The District's policy of only offering a gifted classroom for fourth graders when there is at least one gifted child in the fourth grade cohort provides a design for checking these impacts, and also verifying the impacts on the high achievers themselves.[28] Specifically, consider the top 20 or so fourth grade students (ranked by third grade test scores) at schools with either 0 or a small number (e.g., 1 or 2) of gifted students in the fourth grade cohort. At schools with no gifted children these students should be assigned to regular classrooms, whereas at the schools with a small number of gifted children, the top 20 high achievers should be assigned to a separate gifted classroom. In both sets of schools the students ranked 25-44 should remain pooled with other (lower-ability) students so any effect of the presence of a gifted child on these students is due to a spillover effect.

[28] We are grateful to Kelley Bedard for a suggestion that motivated this section.

To implement this design we identified a set of fourth grade classes at elementary schools in the District with from 0 to 4 gifted children in the entire class.[29] We then identified the students ranked from 1 to 20 on the previous year's achievement scores in each class, and those ranked 25-44. Table 8 presents a series of models for the mean outcomes of the two groups of students. In each case we show the coefficient of a dummy for having at least one gifted child in the (school-wide) class. Columns 1 and 5 present models for third grade (i.e., pre-intervention) average test scores in reading and math. Columns 2 and 6 present models for the fraction of the group that is placed in a gifted classroom. Finally columns 3-4 and 7-8 show models for the group average scores in fourth grade reading and math tests. We present a baseline model with only a linear control for the number of gifted students in the cohort (row 1); then add class-level controls (row 2) and school fixed effects (row 3).[30] Finally, we present results based on a subsample of schools and classes with either zero or one gifted student in the class in row 4.

Looking at the baseline (third grade) test score results in columns 1 and 5, notice that while average student achievement in either the 1-20 rank group or the 25-44 rank group is positively correlated with the presence of at least one gifted child in the cohort (row 1), this correlation is eliminated by the cohort and school-level controls (rows 2 and 3). The first stage models for the probability of assignment to a gifted classroom (columns 2 and 6) show that for the 1-20 rank group, the presence of a gifted child raises the probability of being assigned by 25-33 percentage points, whereas for the 24-44 rank group the corresponding effect is 5-6 percentage points. The <100% impact on the 1-20 group and the positive impact on the 25-44 group are due to our difficulties (noted earlier) in replicating the rankings used by the schools to fill their gifted classes. Despite this slippage, the

[29] This sample comprises 255 fourth grade classes at 94 elementary schools during the years 2009-12 (for which we know the District's ranking formula), and it includes roughly half of all the fourth grade classes during these years.
[30] The class level controls include mean average standardized scores for all students in the class in third grade reading and mathematics, the school-wide FRL rate, the number of students in the class, and the number of students in the class with non-missing test scores in third and fourth grades.

presence of a gifted child has relatively large impact on the probability of placement in a gifted classroom for the top-ranked group, and only a very small effect for their lower-ranked classmates.

Apart from the estimate from the simple specification in row 1 of column 3 (which is presumably upward-biased by about 0.15, given the baseline model in column 1 for this same specification), all the outcome models in columns 3-4 for the 1-20 ranked group show a positive effect of about 0.07 standard deviation units. The implied treatment on the treated effect is in the range of 0.2 to 0.3 standard deviation units - similar to the implied impacts from our individual-level analysis. The estimated effects for the 24-44 ranked group (columns 7-8), in contrast, are all very close to zero and insignificant (apart from the specification in row 1 column 7, which is also upward biased). Importantly, these conclusions hold when we add school dummies to the models (row 3), and therefore derive identification only from cross-cohort comparisons within schools that sometimes have 1-4 gifted children in a class, and sometimes have no gifted children in a class. Moreover, the results are similar when we restrict attention to the subset of school-classes that have either no gifted children, or only one gifted child (row 4). We conclude that moving the top-scoring children in a class to a separate gifted classroom has a positive effect on their reading and math achievement, but little or no effect on the achievement gains of the next highest-scoring group of students in a class.

VI. Interpreting the Differences across Groups

Our analysis suggests that participation in gifted classrooms has different effects on non-gifted high achievers than on Plan A or Plan B gifted students. The comparison between high achievers and the Plan B group is particularly compelling because the two groups have similar levels of prior achievement and similar distributions across schools in the District.[31]

[31] To check that the small differences in shares of the two groups in different schools do not explain the differences, we re-weighted the high achiever sample to have the same distribution across schools as the Plan B

A key difference between Plan B students and high achievers is that the former are selected on the basis of IQ scores whereas the latter are selected on the basis of past achievement. These differences are magnified by our RD estimation strategy, which identifies treatment effects for Plan B "compliers"—students whose assignment to a gifted classroom depends on their IQ score. High-IQ students with strong academic achievement are "always takers" who enter a gifted classroom either as high achievers or as gifted students. The compliers are students who are *not* at the top of their class in terms of their standardized test scores, despite their relatively high cognitive ability – i.e., academic "underachievers".

To provide more systematic evidence on this phenomenon, we used the method proposed by Walters (2013) to derive estimates of the mean past achievement levels and mean cognitive abilities of the compliers in our RD analyses of the Plan A, Plan B, and high achiever groups.[32] Since we lack IQ scores for most high achievers, we use scores on the NNAT screening test as our measure of cognitive ability. The results are summarized in Table 9. We show estimated mean third grade scores in math and reading, mean NNAT scores, and the mean "achievement gap" for each group, which we define as the difference between average standardized achievement scores in math and reading and standardized NNAT scores.

The estimates in Table 9 show, first, that the compliers in our Plan A and Plan B analyses have significantly higher standardized cognitive ability scores than standardized achievement scores. The "achievement gap" for the Plan B compliers is particularly large—around 0.9 standard deviation units. In contrast the achievement scores of the compliers in our high achiever sample are comparable to their cognitive ability scores. Hence, relative to the Plan A and Plan B compliers, the high achievers over-

---

sample and re-estimated our main models for the high achievers. The estimated effects were nearly identical to the effects from unweighted models.

[32] See the Appendix for details.

perform on standardized tests, suggesting the presence of non-cognitive skills that enhance their academic achievement.

A potential explanation for our finding that gifted classrooms raise the math and reading scores of high achievers but not those of the gifted groups is that the teaching methods used in these classes are particularly suited to students with strong non-cognitive skills like motivation and willingness to please the teacher, and are less effective for students who lack these skills. This explanation is consistent with the *drop* in student satisfaction for Plan B students who are assigned to a gifted classroom consisting mostly of high-achievers. It is also consistent with qualitative evidence from interviews with gifted teachers that many of these teachers find it easier to teach high achievers than gifted children.

VII. Conclusions

XXX

References

Abdulkadiroglu, Atila, Joshua D. Angrist and Parag A. Pathak. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82(1): 137-196.

Adelson, Jill L. D. Betsy McCoach and M. Katherine Gavin. "Examining the Effects of Gifted Programming in Mathematics and Reading Using the ECLS-K. *Gifted Child Quarterly* 56(1): pp. 25-39.

Angrist, Joshua D. "The Perils of Peer Effects." NBER Working Paper 19774, December 2013.

Angrist, Joshua D. and Miikka Rokkanen. "Wanna Get Away? RD Identification Away from the Cutoff." NBER Working Paper No. 18662. December 2012.

Bajari, Patrick, Han Hong, Minjung Park and Robert Town. "Regression Discontinuity Designs with an Endogenous Forcing Variable and an Application to Contracting in Health. NBER Working Paper 17643, December 2011.

Barreca, Alan, Jason M. Lindo and Glen R. Waddell. "Heaping Induced Biases in Regression Discontinuity Designs. NBER Working Paper 17408, September 2011.

Betts, Julian R. "The Economics of Tracking in Education." Education." In Hanushek, Eric A., Stephen Machin and Ludger Woessmann (eds.) *Handbook of the Economics of Education*, Volume 3, Amsterdam: North Holland, pp. 341-381.

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

Bhatt, Rachana. "The Impacts of Gifted and Talented Education." Unpublished Working Paper, Department of Economics Georgia State University, August 2012.

Brown, Scott W. Joseph S. Renzulli, E. Jean Gubbins, Del Siegle, Wanli Zhang and Ching-Hui Chen. "Assumptions Underlying the Identification of Gifted Children." Gifted Child Quarterly 40(1) (2005): pp. 68-79.

Bui, Sa A., Steven G. Craig, and Scott A. Imberman. "Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Progams on Achievement." NBER Working Paper 17089, May 2011.

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

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

Davis, Bilie, John Engberg, Dennis N. Epple and Ron Zimmer. "Evaluating the Gifted Program of an Urban District Using a Modified Regression Discontinuity Design." NBER Working Paper 16414, September 2010.

Dobbie, Will and Roland G. Fryer Jr. "Exam High Schools and Academic Achievement: Evidence from New York City." NBER Working Paper 17286, August 2011.

Donovan, M. Suzanne and Christopher T. Cross. *Minority Students in Special and Gifted Education*. Washington DC: National Academies Press, 2002.

Duncan, Greg J. and Richard J. Murnane. *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. New York: Russell Sage Foundation, 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 (August 2011), pp. 1739-1174.

Florida Department of Education. "Standard Error of Measurement". Technical Assistance Paper Number FY 1996-7. Available at http://www.fldoe.org/bii/Gifted_Ed/.

Ford, Donna Y. "The Underrepresentation of Minority Students in Gifted Education: Problems and Promises in Recruitment and Retention." *Journal of Special Education* 32(1) 1998, pp. 4-14.

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

Guyon, Nina, Eric Maurin and Sandra McNally. "The Effect of Tracking Students by Ability into Different Schools: A Natural Experiment." *Journal of Human Resources* 47(3) (2012), pp. 684-721.

Hoxby, Caroline M. and Christopher Avery. "The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students." NBER Working Paper 18586, December 2012.

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

Jolly, Jennifer L. "A Resuscitation of Gifted Education." *American Educational History Journal* 36(1) (2009), pp. 37-52.

Kerr, Sari Pekkala, Tuomas Pekkarinen, and Roope Uuritalo. "School Tracking and Development of Cognitive Skills." *Journal of Labor Economics* 31(3) (July, 2013), pp. 577-602.

Lee, David S. "Randomized Experiments from Non-random Selection in U.S. House Elections," *Journal of Econometrics* 142(2) (February 2008), pp 675–697.

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

Lohman, David F, Katrina A. Korb and Joni Lakin. "Identifying Academically Gifted English-Language Learners Using Nonverbal Tests : A Comparison of the Raven, NNAT, and CogAT." *Gifted Child Quarterly* 52(4) (Fall 2008), pp. 275-296.

Marland, S. P., Jr. *Education of the Gifted and Talented: Report to the Congress of the United States by the U.S. Commissioner of Education.* Washington, DC: U.S. Government Printing Office, 1972.

Mcclain, Mary-Caherine and Seven Pfeiffer. "Identification of Gifted Students in the United States Today: A Look at State Definitions, Policies and Practices." *Journal of Applied School Psychology* 28(1) (2012), pp. 59-88.

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

Pop-Eleches , Cristian and Miguel Urquiola. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review* 103(4) (June 2013), pp. 1289-1324.

Renzulli, Joseph S. (1978). "What Makes Giftedness? Re-examining a Definition." *Phi Delta Kappan* 60(3), pp. 180-184.

Sacerdote, Bruce. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?**"** In Eric A. Hanushek, Stephen Machin and Ludger Woessmann, editors, *Handbook of the Economics of Education* Amsterdam: Elsevier, 2011, Volume 3, pp. 249-277.

Slavin, Robert E. "Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis." *Review of Educational Research* 57(3) (Fall 1987), pp. 293-336.

U.S. Department of Education, Office of Educational Research and Improvement. *National Excellence: A Case for Developing America's Talent*. Washington, DC: U.S. Government Printing Office, 1993.

VanTassel-Baska, Joyce and Elissa F. Brown. "Toward Best Practice: An Analysis of the Efficacy of Curriculum Models in Gifted Education. *Gifted Child Quarterly* 51(4) (Fall 2007), pp.342-358.

Vardardottir, Arna. "Peer Effects and Academic Achievement." *Economics of Education Review* (2013) 36: 108–121.

Walters, Christopher. "A Structural Model of Charter School Choice and Academic Achievement." Unpublished Working Paper, UC Berkeley Department of Economics, September 2013.

Table 1. Sample characteristics

	All students in district for 3rd grade (2004-10)	Plan A, IQ Є [120,140]		Plan B, IQ Є [105,125]		High Achievers, +/- 10 from cutoff	
		In 4th grade gifted class	Not in 4th grade gifted class	In 4th grade gifted class	Not in 4th grade gifted class	In 4th grade gifted class	Not in 4th grade gifted class
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Student demographics							
Female (%)	49%	48%	47%	51%	48%	54%	49%
White (%)	30%	57%	60%	21%	18%	36%	32%
Black (%)	36%	7%	7%	33%	35%	26%	30%
Hispanic (%)	27%	22%	22%	36%	39%	28%	27%
Free lunch eligible (%)	50%	--	--	88%	85%	45%	52%
English language learner (%)	12%	--	--	13%	19%	1%	2%
Plan B eligible (%)	55%	--	--	100%	100%	45%	53%
Median income in ZIP ($1,000s)	57.8	72.3	72.1	51.1	53.8	59.2	59.8
Mean IQ score	104.1	130.5	123.2	117.4	111.1	112.6	108.9
Percent taking test	*21%*	*100%*	*100%*	*100%*	*100%*	*12%*	*10%*
Mean NNAT (screening test)	105.8	130.3	128.4	125.2	122.5	115.1	112.1
Percent taking test	*63%*	*78%*	*74%*	*84%*	*88%*	*67%*	*72%*
Third grade state test scores (standardized)							
Mean reading test	0.07	1.45	1.09	0.91	0.29	1.04	0.73
Mean math test	0.05	1.46	1.13	0.98	0.42	0.88	0.74
Fourth grade state test scores (standardized)							
Mean reading test	0.07	1.32	1.01	0.86	0.27	0.87	0.56
Mean math test	0.07	1.39	1.10	0.91	0.40	0.84	0.58
Mean writing test	0.05	0.83	0.56	0.53	0.12	0.54	0.30
Fourth grade peers' third grade scores (standardized)							
Mean reading test	0.07	1.14	0.31	0.74	-0.06	1.00	-0.03
Mean math test	0.05	1.16	0.32	0.73	-0.06	0.95	-0.02
Number of observations	*138,172*	*1,954*	*725*	*1,872*	*2,600*	*1,709*	*2,435*

Note: Sample in column 1 includes one observation per student for students observed in a non-charter district elementary school in 3rd grade between 2004 and 2011. Third (fourth) grade test scores are reported only for those who stayed in the district through third (fourth) grade and who advanced one grade level per year. Free lunch status (FRL) and English language learner (ELL) status are measured at the end of the first year the student is observed. Sub-samples in columns 2-5 include students who received a first IQ test in 2004-2010 while completing grades 1-3 (if tested in spring) or entering grades 2-4 (if tested in summer/fall), with IQ score within 10 points of the gifted eligibility threshold, and who wrote state standardized tests in both 3rd and 4th grade. Non-verbal ability index (NNAT) is measured at the end of 2nd grade for students who took the test between 2005-2009. Sub-samples in columns 6-7 include students in 4th grade in 2009-2012 in schools that complied with District's ranking formula and had at least one self-contained gifted/high-achiever classroom for 4th graders. State 3rd and 4th grade test scores are in standard deviation units (standardized across district within year and grade.)

Table 2. Regression Discontinuity Estimates for Plan A Gifted Sample

	Baseline Achievement		First Stage Models			Reduced Form Outcomes (Fourth Grade Scores)		
	Baseline Reading	Baseline Math	Probability in Gifted Classroom	Probability Classified as Gifted	Mean Peer Lagged Scores	Reading	Math	Writing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. No controls	0.023	0.088	0.270**	0.498**	0.173**	-0.016	-0.033	0.026
	(0.073)	(0.061)	(0.044)	(0.044)	(0.049)	(0.059)	(0.070)	(0.088)
2. Cohort and school fixed effects	0.033	0.094	0.267**	0.494**	0.143**	-0.052	-0.084	-0.083
and student controls	(0.075)	(0.059)	(0.046)	(0.044)	(0.047)	(0.050)	(0.064)	(0.081)
Sample size	*2679*	*2679*	*2679*	*2679*	*2579*	*2679*	*2679*	*2679*
3. Donut hole specification	0.074	0.100	0.359**	0.609**	0.217**	-0.138	-0.129	-0.112
(exclude IQ scores 127-131)	(0.117)	(0.109)	(0.077)	(0.061)	(0.075)	(0.089)	(0.101)	(0.154)
Sample size	*1976*	*1976*	*1976*	*1976*	*1898*	*1976*	*1976*	*1976*
4. Differenced specification	--	--	--	--	--	-0.022	-0.047	--
(based on change in test scores)						(0.054)	(0.052)	
Sample size	--	--	--	--	--	*3954*	*3954*	--
5. Differenced model with donut ho	--	--	--	--	--	-0.093	-0.034	--
(exclude IQ scores 128-131)						(0.063)	(0.063)	
Sample size	--	--	--	--	--	*3251*	*3251*	--

Notes: Based on Plan A analysis sample described in columns 2-3 of Table 1. Entries are estimated coefficients on a dummy for IQ≥130 from models that include a linear term in IQ interacted with the dummy. Models in rows 2-5 control for student demographics (age, gender, race/ethnicity, and median household income in ZIP code), dummies indicating year of the first IQ test and year in fourth grade, and a complete set of school dummies. Models in rows 2-3, columns 3-8 also control for baseline test scores. Donut hole sample (rows 3 & 5) excludes students with $IQ\in[127,131]$. Differenced models (rows 4-5) are based on expanded bandwidth sample that includes $IQ\in[115,145]$. Standard errors, clustered by school, in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 3. Regression Discontinuity Models for Self-reported Survey Responses

	Plan A Analysis Sample:			Plan B Analysis Sample:			High Achievers Analysis Sample:		
	Mean response (left of threshold)	Local linear RD		Mean response (left of threshold)	Local linear RD		Mean response (left of threshold)	Local linear RD	
		no controls	with controls		no controls	with controls		no controls	with controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. My teacher(s) believe I can succeed	0.793	0.037	0.031	0.833	-0.031	-0.038	0.822	0.001	0.003
		(0.039)	(0.041)		(0.024)	(0.025)		(0.023)	(0.023)
2. My teacher(s) answer questions in a way I understand	0.559	0.014	0.005	0.635	-0.054+	-0.053	0.624	0.004	0.005
		(0.051)	(0.050)		(0.031)	(0.033)		(0.028)	(0.028)
3. I enjoy learning in my school (2007-10 only)	0.383	0.082	0.078	0.598	-0.038	-0.056	0.555	-0.013	-0.013
		(0.068)	(0.069)		(0.034)	(0.035)		(0.036)	(0.036)
4. I feel accepted/like I belong at my school	0.551	0.082+	0.058	0.605	-0.045	-0.051	0.598	-0.013	-0.016
		(0.049)	(0.053)		(0.032)	(0.033)		(0.031)	(0.030)
5. Index (average of nonmissing responses, questions 1-4)	0.592	0.051	0.037	0.665	-0.039+	-0.047*	0.660	-0.004	-0.002
		(0.034)	(0.034)		(0.021)	(0.022)		(0.020)	(0.020)
6. Did not participate in survey	0.051	0.000	-0.003	0.045	0.004	0.000	0.051	-0.002	-0.001
		(0.026)	(0.027)		(0.014)	(0.014)		(0.012)	(0.012)

Notes: Table reports estimated mean responses (for students with IQ or rank just below the threshold) and estimated regression discontinuities in responses (=1 if strongly agreed with statement, 0 otherwise) to survey questions administered at the end of fourth grade by the District. Standard errors, clustered by school, in parentheses.

Table 4. Regression Discontinuity Estimates for Plan B Gifted Sample

	Baseline Achievement		First Stage Models			Reduced Form Outcomes (Fourth Grade Scores)		
	Baseline Reading	Baseline Math	Probability in Gifted Classroom	Probability Classified as Gifted	Mean Peer Lagged Scores	Reading	Math	Writing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. No controls	-0.040	-0.062	0.365**	0.522**	0.200**	-0.028	-0.033	0.118*
	(0.044)	(0.041)	(0.029)	(0.028)	(0.038)	(0.046)	(0.038)	(0.052)
2. Cohort and school fixed effects and student controls	-0.039	-0.045	0.359**	0.519**	0.240**	0.013	0.004	0.116*
	(0.045)	(0.041)	(0.027)	(0.027)	(0.030)	(0.032)	(0.035)	(0.047)
Sample size	*4472*	*4472*	*4472*	*4472*	*4251*	*4472*	*4472*	*4472*
3. Donut hole specification (exclude IQ scores 114-117)	-0.040	-0.030	0.453**	0.612**	0.350**	0.052	0.063	0.179*
	(0.070)	(0.068)	(0.037)	(0.035)	(0.041)	(0.052)	(0.058)	(0.071)
Sample size	*3582*	*3582*	*3582*	*3582*	*3405*	*3582*	*3582*	*3582*
4. Differenced specification (based on change in test scores)	--	--	--	--	--	0.015	-0.005	--
						(0.030)	(0.033)	
Sample size						*6154*	*6154*	
5. Differenced model with donut h (exclude IQ scores 114-117)	--	--	--	--	--	0.038	0.018	--
						(0.048)	(0.043)	
Sample size						*5264*	*5264*	

Notes: Based on Plan B analysis sample described in columns 4-5 of Table 1. Entries are estimated coefficients on a dummy for IQ≥116 from models that include a linear term in IQ interacted with the dummy. Models in rows 2-5 control for student demographics (age, gender, race/ethnicity, and median household income in ZIP code), dummies indicating year of the first IQ test and year in fourth grade, and a complete set of school dummies. Models in rows 2-3, columns 3-8 also control for baseline test scores. Donut hole sample (rows 3 & 5) excludes students with IQ∈[114,117]. Value added models (rows 4-5) based on expanded bandwidth sample that includes IQ∈[100,130]. Standard errors, clustered by school, in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 5. Plan B Heterogeneity Analysis

A. Interaction effects for fourth grade achievement

	Baseline Achievement		First Stage Models			Reduced Form Outcomes (4th/8th Grade)		
	Baseline Reading (1)	Baseline Math (2)	Probability in Gifted Classroom (3)	Probability Classified as Gifted (4)	Mean Peer Lagged Scores (5)	Reading (6)	Math (7)	Writing (8)
Interaction variable:								
1. early cohorts (4th grade in 2005-2008)								
main effect (later cohorts)	-0.035	-0.058	0.337**	0.539**	0.236**	0.055	0.012	0.037
	(0.058)	(0.058)	(0.040)	(0.039)	(0.047)	(0.042)	(0.053)	(0.068)
interaction effect (early cohorts)	-0.011	0.009	0.052	-0.045	0.001	-0.089	-0.009	0.182
	(0.089)	(0.076)	(0.059)	(0.049)	(0.066)	(0.067)	(0.071)	(0.113)
2. gender (boys)								
main effect (girls)	0.011	-0.007	0.332**	0.513**	0.219**	-0.009	-0.013	0.053
	(0.067)	(0.059)	(0.040)	(0.034)	(0.046)	(0.045)	(0.047)	(0.071)
interaction effect (boys)	-0.098	-0.087	0.053	0.011	0.040	0.043	0.034	0.125
	(0.102)	(0.090)	(0.049)	(0.039)	(0.066)	(0.065)	(0.071)	(0.108)
3. schools with high % free lunch (> sample median)								
main effect (low FRL rate)	-0.031	-0.040	0.296**	0.457**	0.260**	0.018	0.014	-0.019
	(0.071)	(0.059)	(0.034)	(0.034)	(0.040)	(0.043)	(0.051)	(0.073)
interaction effect (high FRL rate)	-0.019	-0.022	0.122*	0.120*	-0.041	-0.010	-0.019	0.262*
	(0.095)	(0.084)	(0.051)	(0.048)	(0.060)	(0.063)	(0.068)	(0.108)
4. schools with low average peer quality in gifted class (< sample median)								
main effect (high quality peers)	0.025	-0.045	0.317**	0.452**	0.322**	0.012	0.030	0.035
	(0.062)	(0.056)	(0.034)	(0.035)	(0.040)	(0.039)	(0.046)	(0.065)
interaction effect (low quality)	-0.129	-0.009	0.082	0.133**	-0.154**	0.004	-0.051	0.170+
	(0.085)	(0.081)	(0.056)	(0.045)	(0.057)	(0.061)	(0.060)	(0.095)
number of observations	*4472*	*4472*	*4472*	*4472*	*4251*	*4472*	*4472*	*4472*

Note: Table continues. See notes at end of second page.

Table 5. Plan B Heterogeneity Analysis , continued

B. Early cohorts only: fourth grade vs. eighth grade achievement with gender interaction

	Baseline Achievement		First Stage Models (4th Grade Status)			Reduced Form Outcomes (4th/8th Grade)		
	Baseline Reading	Baseline Math	Probability in Gifted Classroom	Probability Classified as Gifted	Mean Peer Lagged Scores	Reading	Math	Writing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
5. fourth grade outcomes for early cohort only, with gender interaction								
main effect (girls)	-0.017	0.061	0.384**	0.503**	0.279**	-0.004	-0.004	0.150
	(0.096)	(0.076)	(0.059)	(0.047)	(0.069)	(0.071)	(0.070)	(0.098)
interaction effect (boys)	-0.056	-0.172	0.006	-0.004	-0.112	-0.043	0.020	0.182
	(0.148)	(0.129)	(0.071)	(0.055)	(0.092)	(0.099)	(0.113)	(0.146)
number of observations	*2181*	*2181*	*2181*	*2181*	*2081*	*2181*	*2181*	*2181*
6. eighth grade outcomes for early cohort only, with gender interaction								
main effect (girls)	-0.070	0.059	0.372**	0.510**	0.198**	-0.068	-0.030	-0.143
	(0.110)	(0.090)	(0.064)	(0.051)	(0.071)	(0.095)	(0.090)	(0.127)
interaction effect (boys)	-0.011	-0.126	0.043	0.033	0.007	0.054	-0.008	0.368*
	(0.165)	(0.147)	(0.076)	(0.061)	(0.096)	(0.125)	(0.122)	(0.165)
number of observations	*1703*	*1703*	*1703*	*1703*	*1623*	*1703*	*1703*	*1703*

Notes: see notes to table 4. RD models include main effect (dummy for IQ score above 116 point threshold) and interaction of RD main effect with specified indicator.   Models in panel A are fit to Plan B analysis sample. Models in panel B are fit to observations from early cohorts in Plan B analysis sample. First stage models for all specifications refer to fourth grade. Reduced form outcomes refer to 4th grade for all models except specification in row 6.

Table 6a. Regression Discontinuity Estimates for High Achiever Sample

	Baseline Achievement		First Stage Models		Reduced Form Outcomes (Fourth Grade Scores)		
	Baseline Reading	Baseline Math	Probability in Gifted Classroom	Mean Peer Lagged Scores	Reading	Math	Writing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. No controls	0.008	-0.046	0.323**	0.273**	0.092**	0.073+	-0.011
	(0.029)	(0.043)	(0.025)	(0.028)	(0.034)	(0.039)	(0.054)
2. Cohort and school fixed effects and student controls	0.015	-0.044	0.319**	0.277**	0.093**	0.087*	-0.012
	(0.027)	(0.040)	(0.026)	(0.029)	(0.031)	(0.035)	(0.051)
3. Differenced specification (based on change in test scores)	--	--	--	--	0.092**	0.105*	--
					(0.033)	(0.041)	
Sample size	*4144*	*4144*	*4144*	*4041*	*4144*	*4144*	*4144*

Notes: Based on high achiever analysis sample, described in columns 6-7 of Table 1. Standard errors, clustered by school, in parentheses. See text for more details of model specifications.

Table 6b. Regression Discontinuity Estimates for 5th Grade Outcomes of Fourth Grade High Achiever Sample

	Probability Stayed in District through 5th Grade	Probability Placed in 5th Grade Gifted Classroom	Baseline Achievement (3rd Grade)		Reduced Form Outcomes (Fifth Grade Scores)		
			Baseline Reading	Baseline Math	5th Grade Reading	5th Grade Math	5th Grade Science
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. No controls	-0.001	0.070**	0.054	-0.049	0.046	0.078+	0.096+
	(0.019)	(0.025)	(0.033)	(0.057)	(0.046)	(0.046)	(0.054)
2. Cohort and school fixed effects and student controls	0.001	0.068*	0.050+	-0.064	0.052	0.077+	.113*
	(0.019)	(0.027)	(0.029)	(0.053)	(0.044)	(0.045)	(0.045)
3. Differenced specification (based on change in test scores)	--	--	--	--	0.013	0.117**	--
					(0.044)	(0.043)	
Sample size	*3089*	*2770*	*2770*	*2770*	*2770*	*2770*	*2037*

Notes: Based on students in high achiever analysis sample who are observed in fourth grade up to 2011. Column 1 includes all students. Columns 2-6 includes only students who are observed though the end of fifth grade. Standard errors, clustered by school, in parentheses. See text for more details of model specifications.

Table 7. High Achiever Heterogeneity Analysis

	Baseline Achievement		First Stage Models		Reduced form outcomes (fourth grade scores)				
	Baseline Reading (1)	Baseline Math (2)	Probability in Gifted Classroom (3)	Mean Peer Lagged Scores (4)	Reading (5)	Reading (Differenced) (6)	Math (7)	Math (Differenced) (8)	Writing (9)
Interaction variable:									
1. gender									
main effect (girls)	0.055	-0.030	0.332**	0.265**	0.037	0.004	0.049	0.080	-0.031
	(0.036)	(0.048)	(0.036)	(0.041)	(0.049)	(0.058)	(0.050)	(0.055)	(0.069)
interaction effect (boys)	-0.082+	-0.022	-0.026	0.024	0.114	0.152+	0.076	0.065	0.034
	(0.045)	(0.058)	(0.050)	(0.062)	(0.070)	(0.080)	(0.077)	(0.084)	(0.101)
2. FRL status									
main effect (not FRL eligible)	0.059+	-0.047	0.376**	0.367**	0.071	0.029	0.075	0.107	0.003
	(0.035)	(0.061)	(0.030)	(0.037)	(0.047)	(0.055)	(0.052)	(0.067)	(0.071)
interaction effect (FRL eligible)	-0.090*	0.009	-0.119*	-0.182**	0.048	0.100	0.021	0.005	-0.031
	(0.040)	(0.072)	(0.056)	(0.062)	(0.062)	(0.070)	(0.065)	(0.080)	(0.105)
3. schools with high % free lunch (> sample median)									
main effect (low FRL rate)	0.036	-0.115+	0.333**	0.284**	0.044	0.006	0.097+	0.168*	-0.013
	(0.040)	(0.068)	(0.030)	(0.036)	(0.045)	(0.054)	(0.054)	(0.071)	(0.070)
interaction effect (high FRL rate)	-0.043	0.149+	-0.027	-0.013	0.097	0.144+	-0.020	-0.111	0.001
	(0.052)	(0.081)	(0.050)	(0.057)	(0.063)	(0.077)	(0.072)	(0.087)	(0.104)
4. schools with low average peer quality in gifted class (< sample median)									
main effect (high peer quality)	0.032	-0.109	0.317**	0.332**	0.045	0.009	0.061	0.128+	-0.018
	(0.042)	(0.068)	(0.037)	(0.044)	(0.046)	(0.059)	(0.056)	(0.072)	(0.064)
interaction effect (low quality)	-0.035	0.138+	0.006	-0.109+	0.098+	0.140+	0.053	-0.031	0.013
	(0.053)	(0.082)	(0.048)	(0.059)	(0.055)	(0.072)	(0.070)	(0.085)	(0.105)
number of observations	*4144*	*4144*	*4144*	*4041*	*4144*	*4144*	*4144*	*4144*	*4144*

Notes: see notes to table 6. RD models include main effect (dummy for IQ score above 116 point threshold) and interaction of RD main effect with specified indicator. All models are fit to high achiever B analysis sample.

Table 8. Estimated group-level models for effect of having at least one student in class designated as gifted in fourth grade

	Students ranked 1-20				Students ranked 25-44			
	Baseline Scores: 3rd grade (reading and math combined)	First Stage: Fraction in gifted classroom in fourth grade	Outcome: 4th grade (reading and math combined)	Outcome: Difference in scores (4th-3rd grade)	Baseline Scores: 3rd grade (reading and math combined)	First Stage: Fraction in gifted classroom in fourth grade	Outcome: 4th grade (reading and math combined)	Outcome: Difference in scores (4th-3rd grade)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<u>Estimation sample: schools/classes with ≤4 gifted students</u>								
1. Controls=linear control for number of gifted students in class	0.148+	0.352**	0.211**	0.064+	0.249*	0.071**	0.239*	-0.002
	(0.078)	(0.062)	(0.077)	(0.032)	(0.095)	(0.025)	(0.093)	(0.042)
2. add class-level controls & year dummies	-0.012	0.330**	0.077+	0.089**	0.047	0.063*	0.047	0.006
	(0.032)	(0.063)	(0.040)	(0.033)	(0.028)	(0.024)	(0.039)	(0.040)
3. add school fixed effects	-0.006	0.241*	0.054	0.061	0.005	0.053	-0.026	-0.018
	(0.039)	(0.094)	(0.062)	(0.053)	(0.030)	(0.034)	(0.055)	(0.057)
number of fourth grade school/cohorts	*255*	*255*	*255*	*255*	*255*	*255*	*255*	*255*
<u>Estimation sample: schools/classes with ≤1 gifted student</u>								
4. class-level controls & year dummies	-0.024	0.408**	0.044	0.067*	0.017	0.083**	0.027	0.012
	(0.029)	(0.057)	(0.039)	(0.031)	(0.028)	(0.016)	(0.033)	(0.037)
number of fourth grade school/cohorts	*116*	*116*	*116*	*116*	*116*	*116*	*116*	*116*

Note: Entries are coefficients on a dummy variable equal to one if the class (i.e., school-wide cohort in a given grade in a given year) had at least one gifted student in fourth grade. Full estimation sample includes 255 fourth grade cohorts at 94 elementary schools during the years 2009-2012. Models in columns 1-4 use average outcomes for top 20 ranked students in class (school-year cohort); models in columns 5-8 use average outcomes for students ranked 25-44 in class. Dependent variables are: group average of reading and math z-scores in 3rd grade (columns 1,5); fraction of students in the group who are in a gifted/high-achiever class in 4th grade (columns 2,5); group average of reading and math z-scores in 4th grade (columns 3,7); and the group average change in reading and math z-scores from 3rd to 4th grade (columns 4, 8). Student rank is based on the formula used by the district in 2009-2012 to determine placement of non-gifted students into 4th grade gifted classrooms. Group averages include gifted students whose rank falls within the indicated range, and excludes students who are missing test scores from either 3rd or 4th grade. Class-level controls include class average 3rd grade math and reading scores; fraction of students who are free lunch eligible; number of students in the class, and the number of students with non-missing test scores in both 3rd and 4th grade. Standard errors in parentheses.

Table 9. Complier characteristics and regression discontinuity estimates for reweighted samples

	Average Complier Characteristics			First Stage Models			Reduced Form Outcomes (Fourth Grade Scores)			
	Baseline Reading (1)	Baseline Math (2)	Achievement-Ability Gap (Mean baseline score - NNAT) (3)	Probability in Gifted Classroom (4)	Probability Classified as Gifted (5)	Mean Peer Lagged Scores (6)	Reading (7)	Math (8)	Writing (9)	Satisfaction Index (10)
Plan A gifted sample										
1. unweighted (as in Table 2, row 2)	**1.31**	**1.32**	**-0.75**	**0.267** (0.046)**	**0.494** (0.044)**	**0.143** (0.047)**	**-0.052 (0.050)**	**-0.084 (0.064)**	**-0.083 (0.081)**	**0.037 (0.034)**
2. reweighted to match Plan B gifted	1.20	1.14	-1.05	0.214** (0.067)	0.464** (0.068)	0.065 (0.069)	-0.231+ (0.126)	-0.155 (0.105)	-0.335* (0.155)	0.032 (0.050)
3. reweighted to match High Achievers	1.44	1.20	-0.83	0.255** (0.067)	0.508** (0.058)	0.092 (0.062)	-0.251* (0.118)	-0.119 (0.096)	-0.262* (0.127)	0.065 (0.050)
Plan B gifted sample										
4. reweighted to match Plan A gifted	0.75	0.88	-0.82	0.241** (0.043)	0.384** (0.036)	0.280** (0.052)	0.083 (0.058)	0.030 (0.062)	0.047 (0.103)	-0.008 (0.056)
5. unweighted (as in Table 4, row 2)	**0.70**	**0.78**	**-0.91**	**0.359** (0.027)**	**0.519** (0.027)**	**0.240** (0.030)**	**0.013 (0.032)**	**0.004 (0.035)**	**0.116* (0.047)**	**-0.047* (0.022)**
6. reweighted to match High Achievers	0.72	0.82	-0.84	0.327** (0.040)	0.532** (0.035)	0.254** (0.045)	0.053 (0.054)	0.012 (0.046)	0.141* (0.068)	-0.110** (0.035)
High Achiever sample										
7. reweighted to match Plan A gifted	1.14	1.09	-0.02	0.404** (0.042)	--	0.320** (0.045)	0.040 (0.042)	0.034 (0.054)	-0.020 (0.085)	0.026 (0.038)
8. reweighted to match Plan B gifted	0.85	0.77	-0.06	0.346** (0.038)	--	0.276** (0.040)	0.116** (0.033)	0.079 (0.051)	0.007 (0.067)	-0.012 (0.028)
9. unweighted (as in Table 6, row 2)	**0.87**	**0.79**	**-0.08**	**0.319** (0.026)**	**--**	**0.277** (0.029)**	**0.093** (0.031)**	**0.087* (0.035)**	**-0.012 (0.051)**	**-0.002 (0.020)**

Notes: Columns 1-2 contain estimated mean baseline achievement test scores for compliers in the RD models. Column 3 shows the estimated mean gap between standardized achievement score and standardized NNAT score. Columns 4-10, rows 1, 5, and 9 reproduce the RD estimates from the models for Plan A, Plan B, and High Achiever groups seen in row 2 of Tables 2, 4, and 6 respectively. The remaining rows in columns 4-10 are RD estimates from the same models using samples that are reweighted to resemble the distribution across schools of each of the other groups. All estimates are from RD models with a full set of controls.

Appendix Table 1. Differences in Characteristics of Students Just Above and Just Below IQ Threshold and Estimated Coefficients from Conditional Logit Model of Probability of Being Above Threshold

	Plan A Sample (127-131 IQ)		Plan B Sample (114-117 IQ)	
	Difference in Means Above-Below (1)	Coefficients from Logit for Pr(Above) (2)	Difference in Means Above-Below (3)	Coefficients from Logit for Pr(Above) (4)
White	0.05	0.33	0.02	0.34
	(0.05)	(0.29)	(0.03)	(0.28)
Black	0.03	0.75	0.04	0.46+
	(0.03)	(0.48)	(0.03)	(0.27)
Hispanic	-0.05	0.03	-0.03	0.24
	(0.04)	(0.32)	(0.03)	(0.26)
Female	-0.01	-0.02	-0.01	-0.07
	(0.05)	(0.20)	(0.04)	(0.15)
Median Income (1000's)	0.25	-0.01	0.56	0.00
	(2.16)	(0.01)	(1.27)	(0.01)
On Free/Reduced Lunch	--	--	0.03	0.28
			(0.03)	(0.20)
School Fraction FRL	-0.02	-0.99	-0.01	-0.17
	(0.02)	(0.69)	(0.02)	(0.39)
NNAT Score (normalized)	-2.57	-0.00	-2.31	-0.00
	(5.37)	(0.00)	(3.16)	(0.00)
3rd grade FCAT Reading (Normalized)	-0.01	-0.12	0.04	0.10
	(0.07)	(0.14)	(0.05)	(0.11)
3rd grade FCAT Math (Normalized)	0.11+	0.33+	0.01	-0.02
	(0.06)	(0.17)	(0.05)	(0.12)
SAT Reading (Normalized)*	0.03	--	0.09	--
	(0.07)		(0.06)	
SAT Math (Normalized)*	0.13+	--	0.02	--
	(0.08)		(0.07)	
P-value for Chi-square test: all coefficients=0	--	0.38		0.73
Number Obs.	703		890	

Notes: Columns 1 and 3 show differences in mean characteristics between students with IQ's just above versus just below the Plan A or Plan B eligibility thresholds. Columns 2 and 4 show estimated coefficients for logistic regression model of probability of being above the threshold, conditional on being just above or just below the threshold. Standard errors in parentheses. * SAT Reading Reading and SAT Math scores are from the year the student first took the IQ test (i.e. the year before potential entry to the gifted program). SAT Reading scores are available only from 2004-2008; SAT Math scores are available only from 2004-2007. Sample sizes for these variables are 610 (Plan A, SAT Reading); 560 (Plan A, SAT Math); 742 (Plan B, SAT Reading); 626 (Plan B, SAT Math).

Figure 1. Fraction of third graders who are classified as gifted by the end of the school year

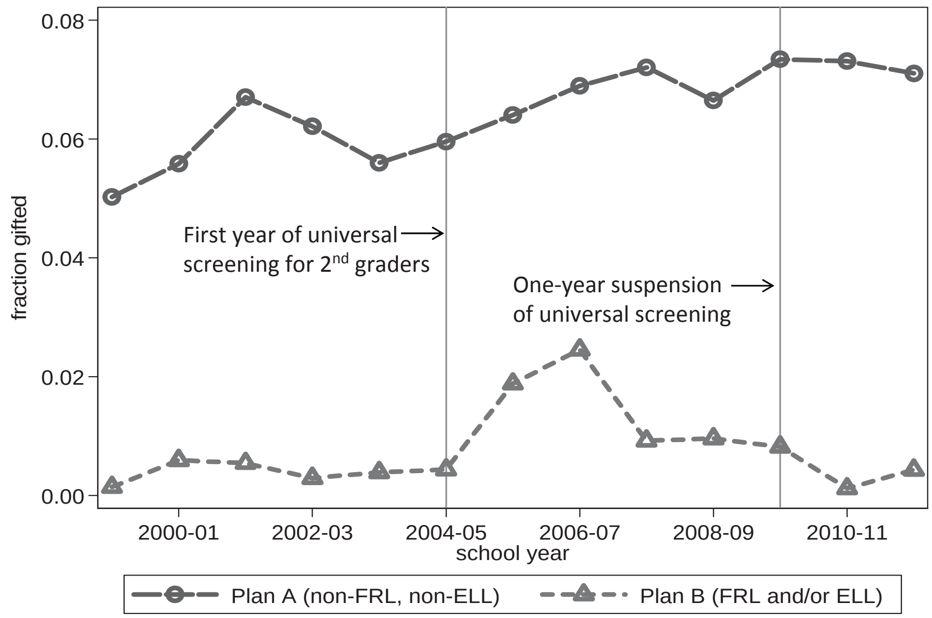


Figure 2. Composition of fourth grade gifted classrooms by %FRL in school

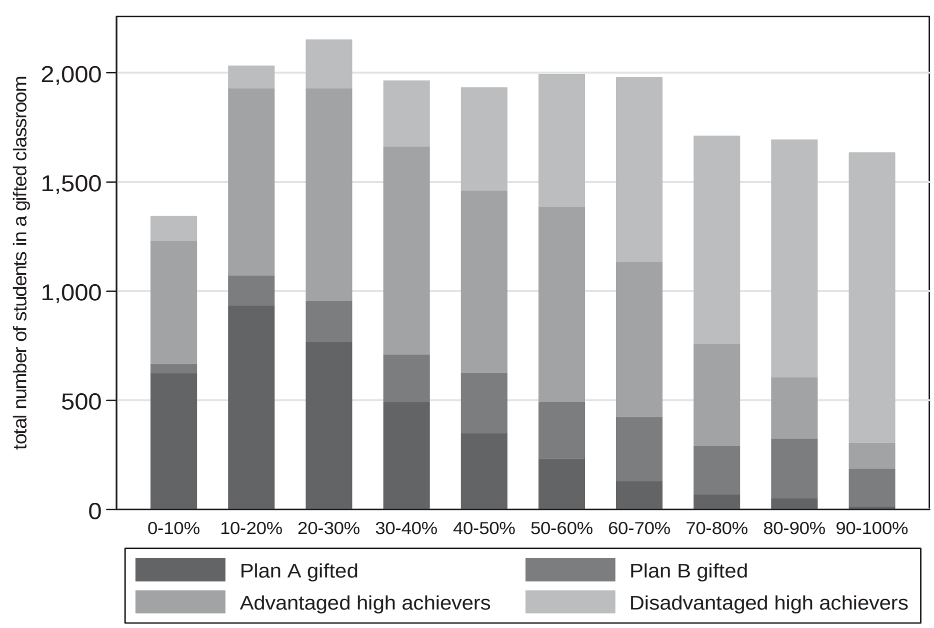
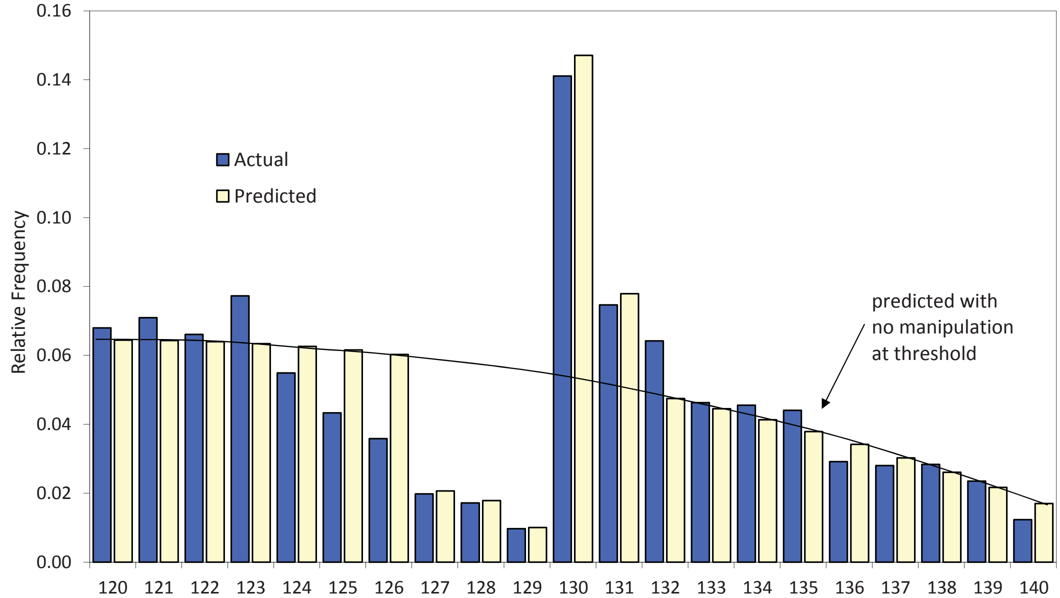
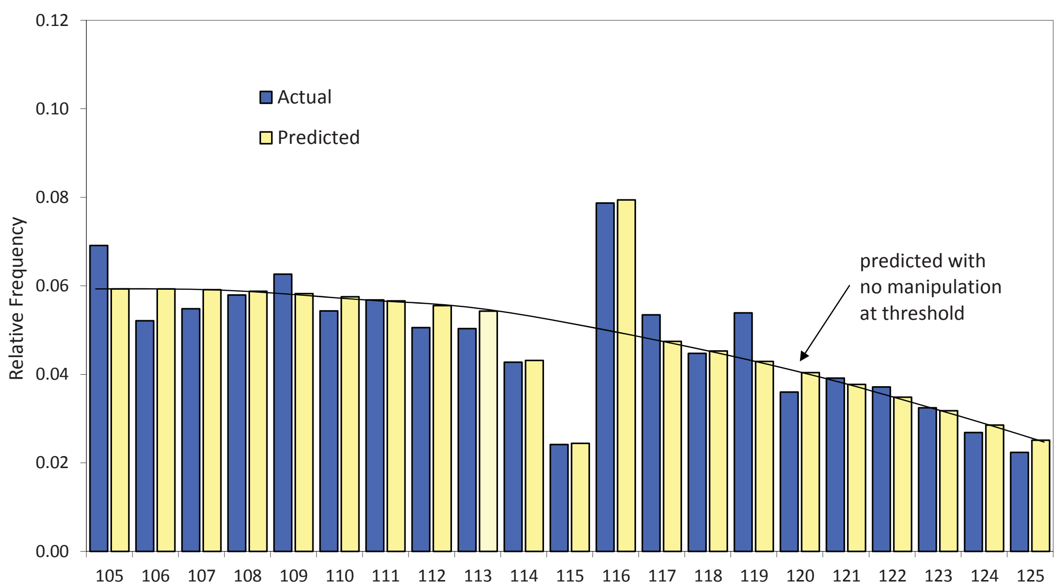


Figure 3a. Histogram of First IQ Scores, Plan A Sample



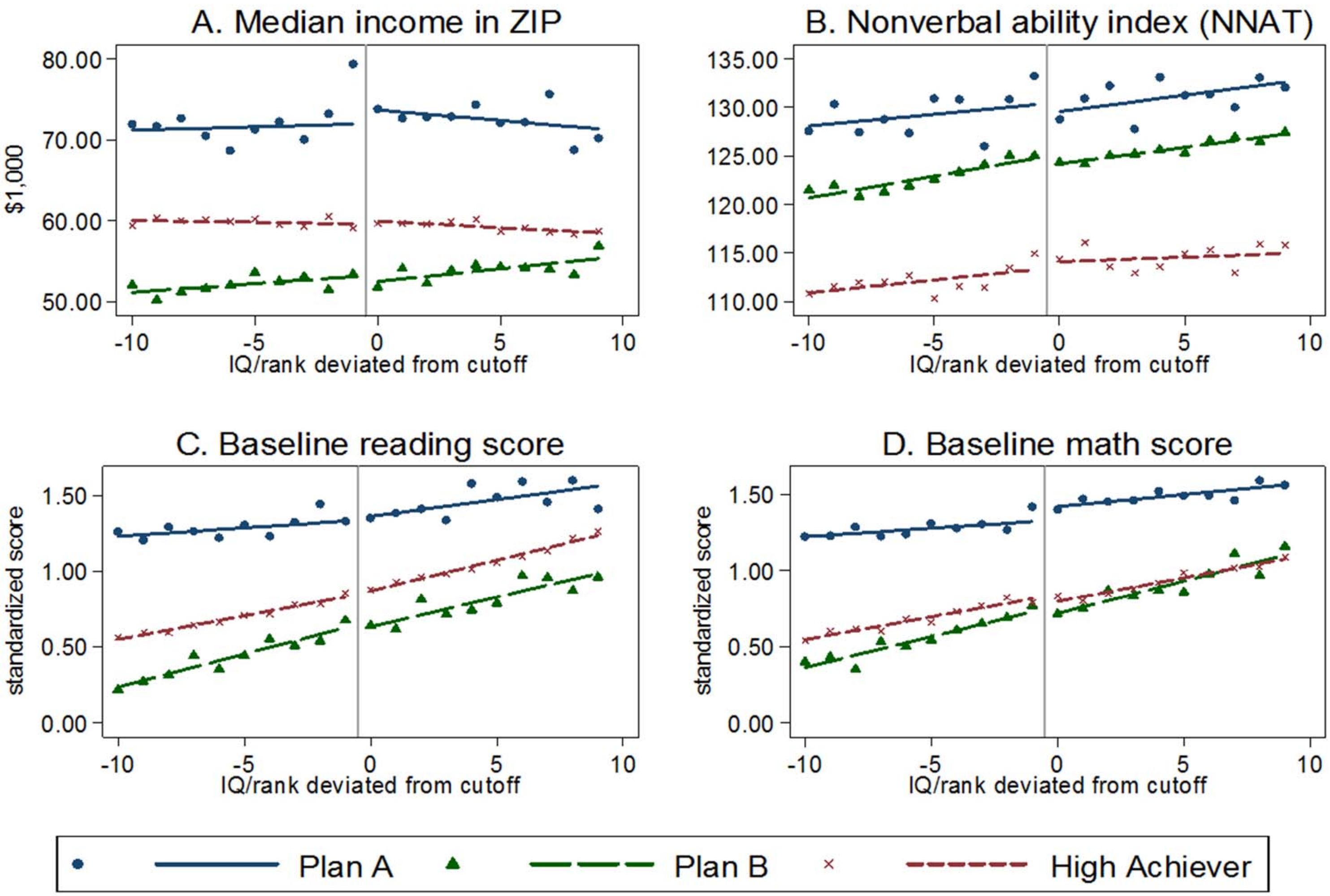
Note: predicted model assumes quadratic, with arbitrary fractions of scores from 127-129 points shifted to 130/131 points. Goodness of fit statistic for model is 82.2 with 14 degrees of freedom. Sample size is 2,679.

Figure 3b. Histogram of First IQ Scores, Plan B Sample



Note: predicted model assumes quadratic, with arbitrary fractions of scores from 114-115 points shifted to 116/117 points. Goodness of fit statistic for model is 35.7 with 15 degrees of freedom. Sample size is 4,472.

Figure 4. Student characteristics and baseline scores, by group



Note: Means and fitted values from local linear regressions. Reading and math scores are standardized within district and year.

Figure 5. First stage relationships

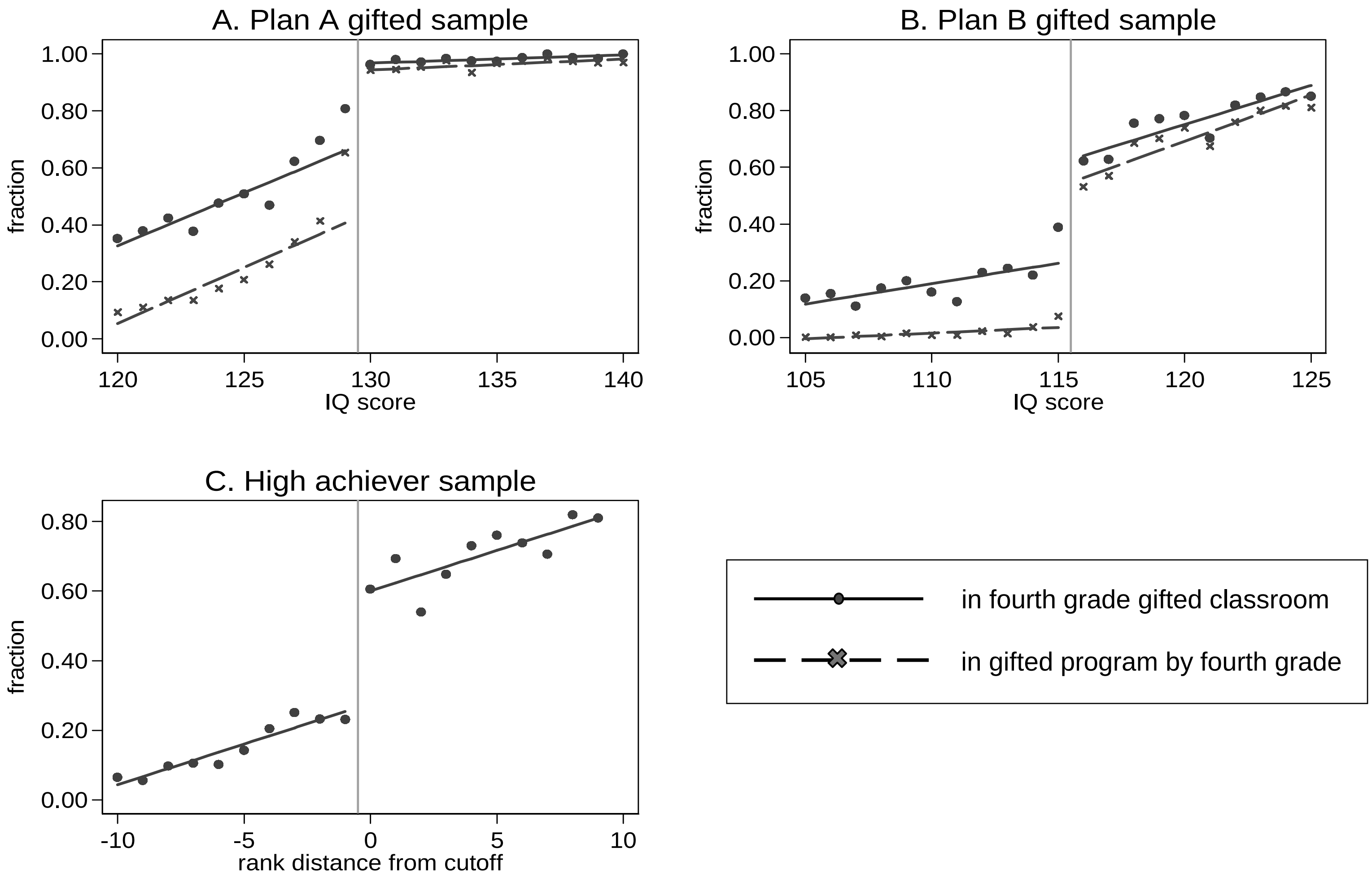
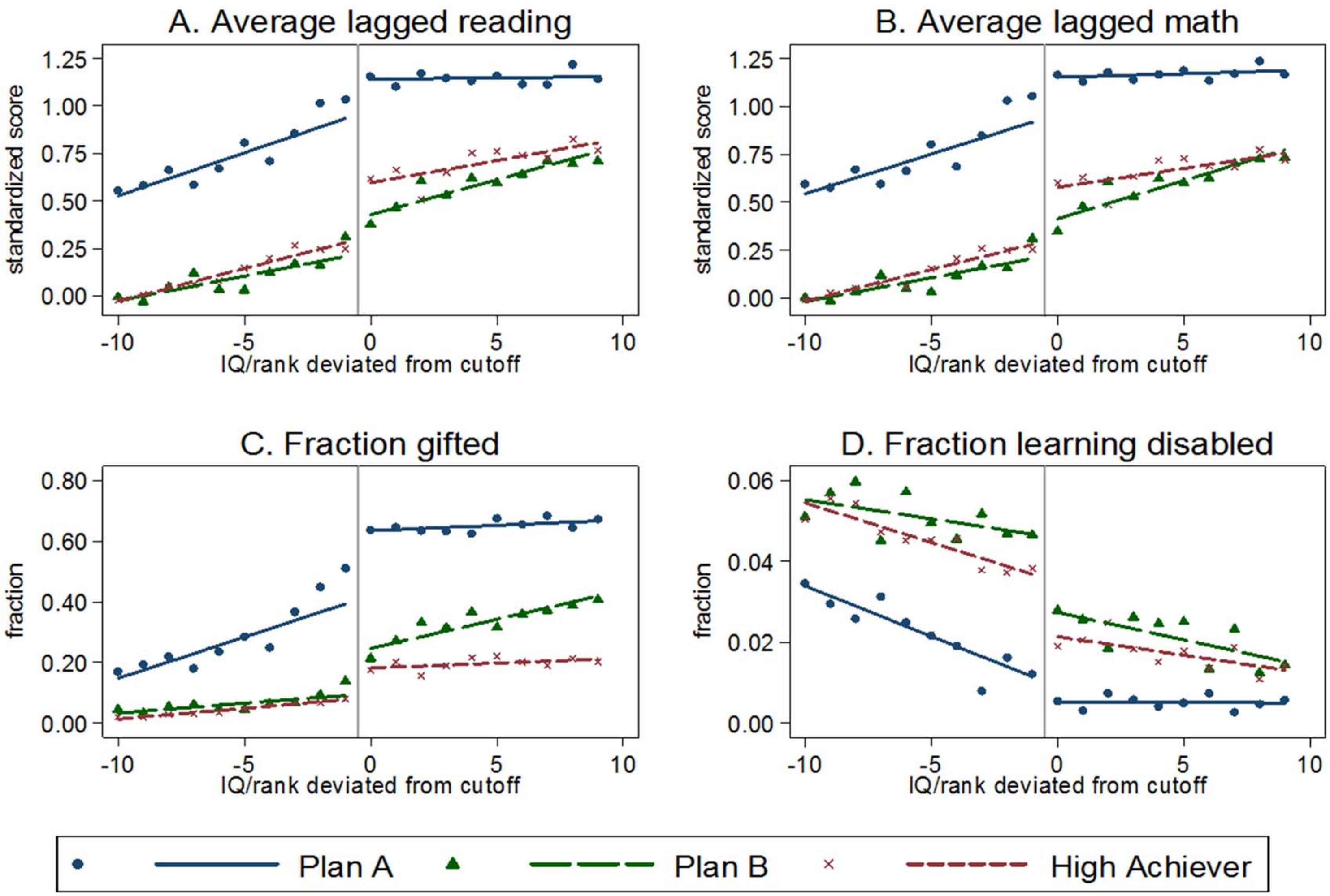


Figure 6. Characteristics of fourth grade classmates, by group



Note: Means and fitted values from local linear regressions. Reading and math scores are standardized within district and year.

Figure 7. Fourth grade standardized test scores, Plan A sample

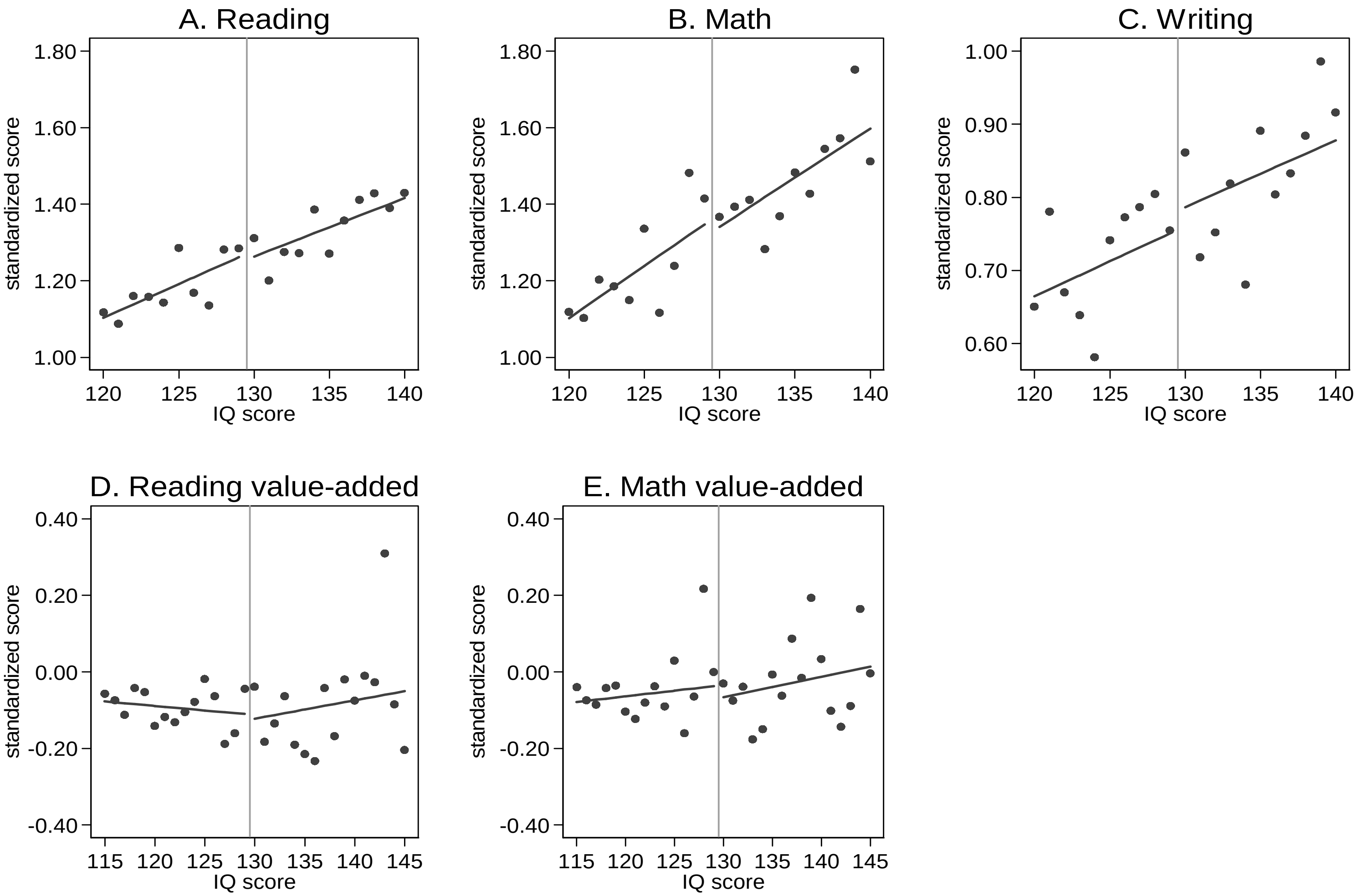


Figure 8. Fourth grade standardized test scores, Plan B sample

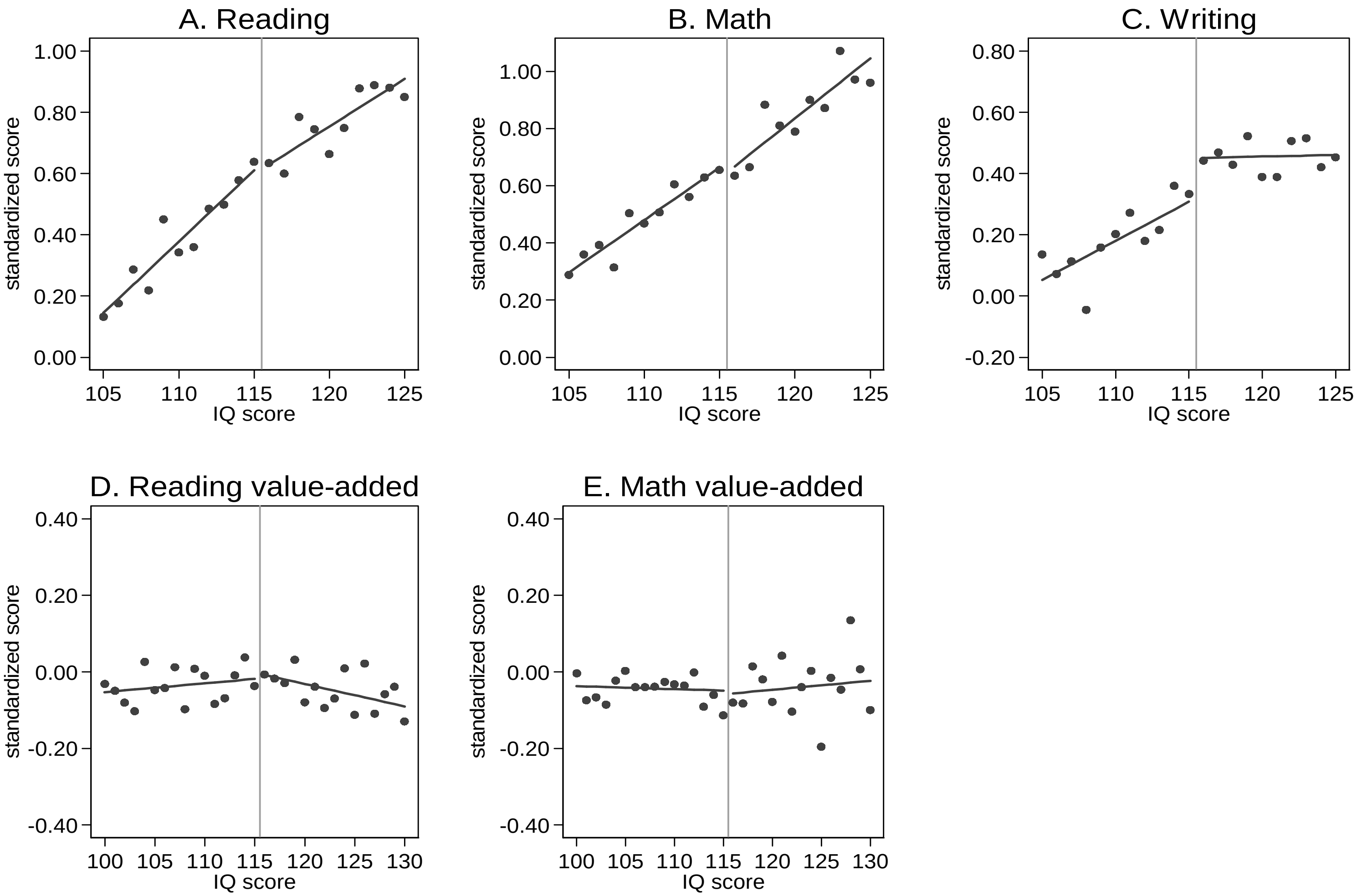


Figure 9. Plan B test scores by gender, grade and cohort

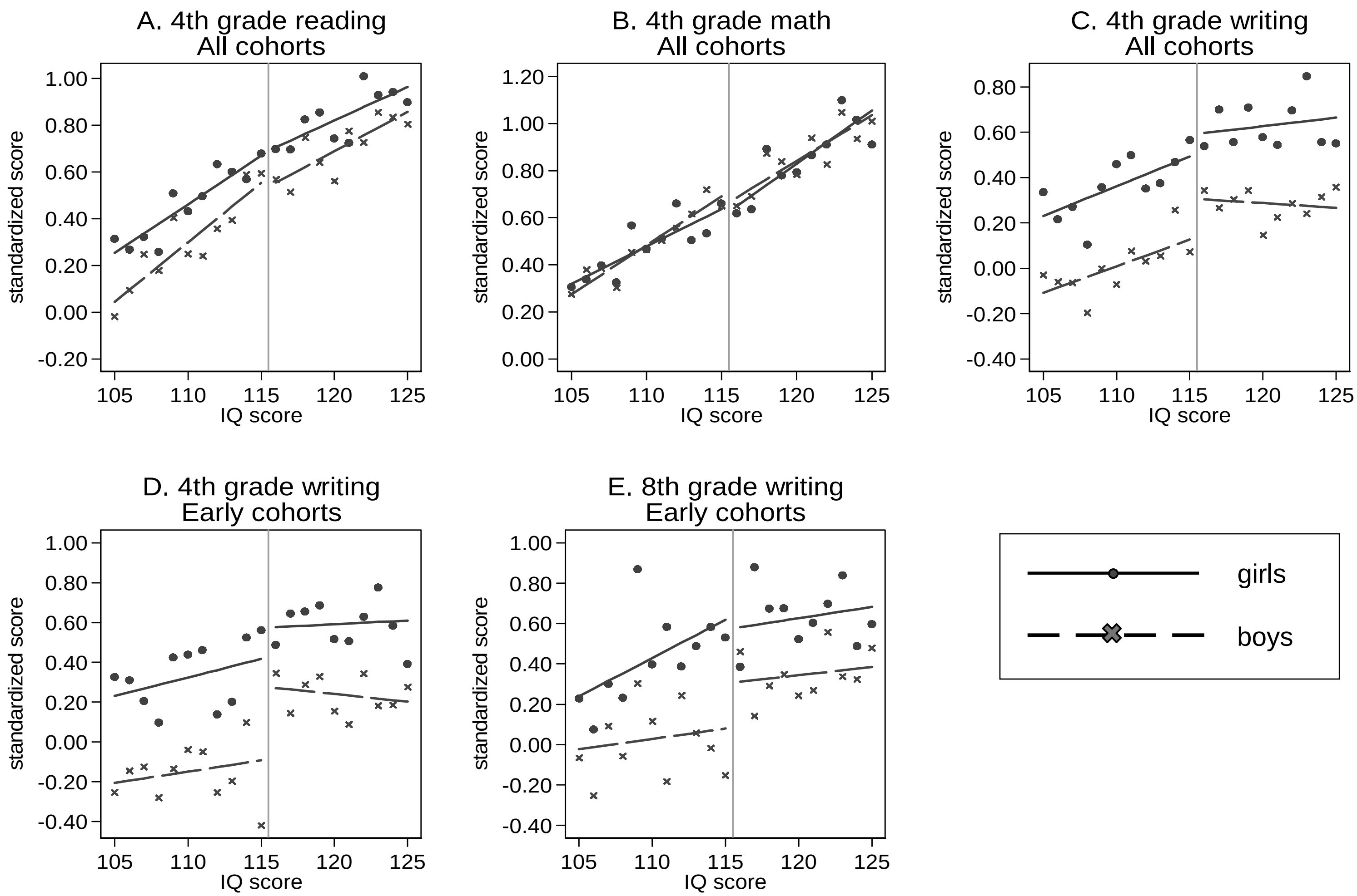
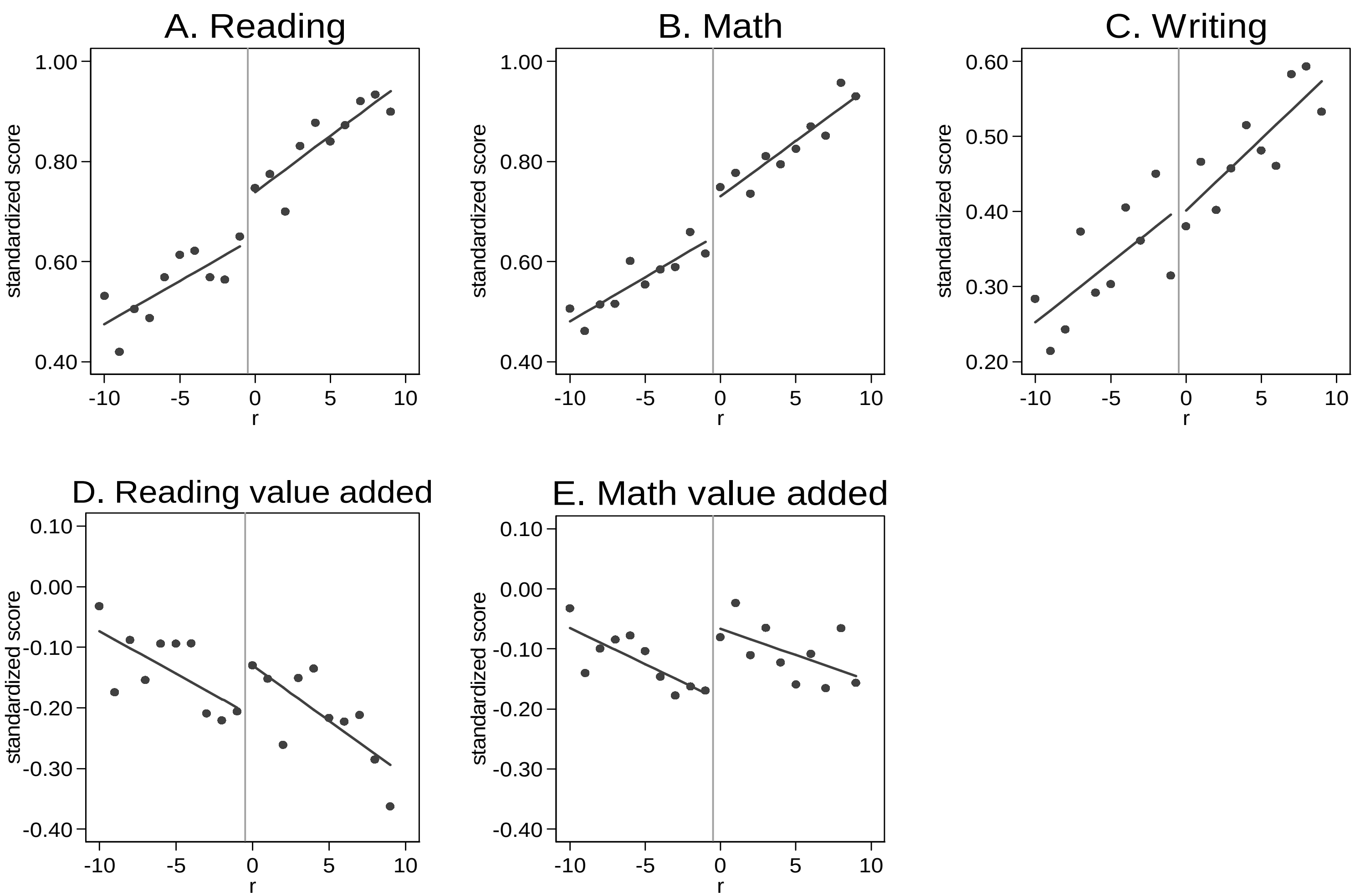
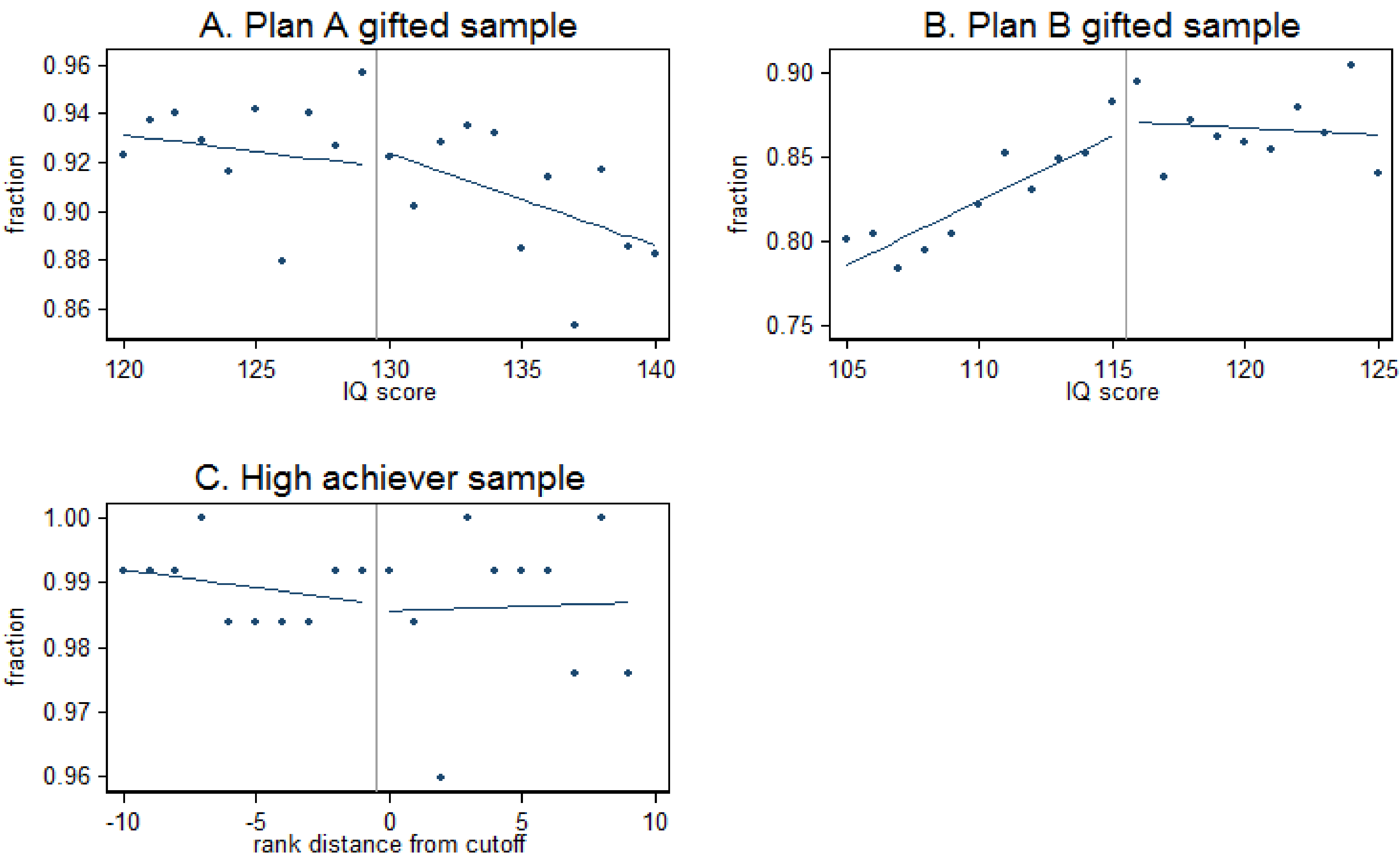


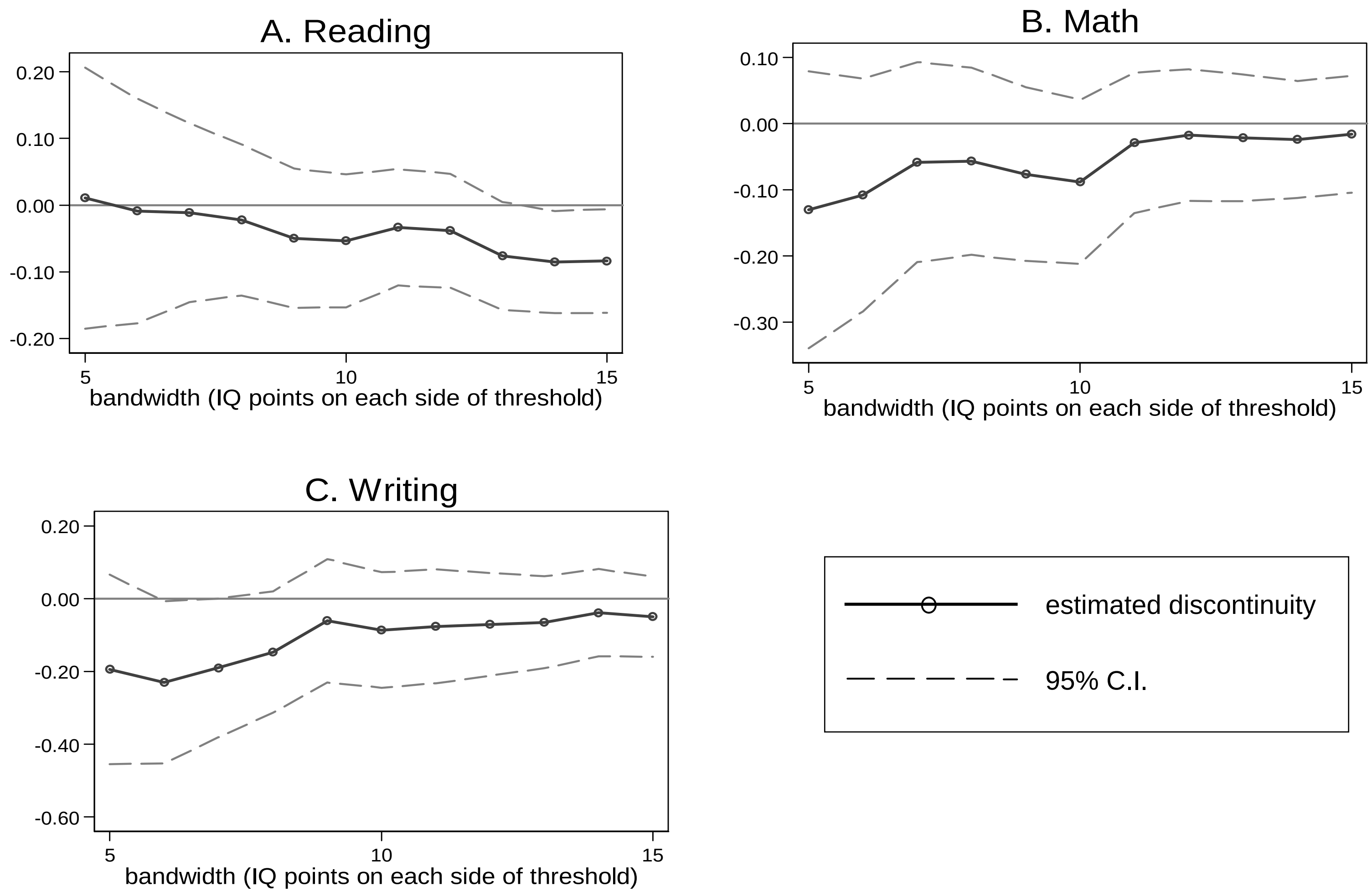
Figure 10. Fourth grade standardized test scores, High achiever sample



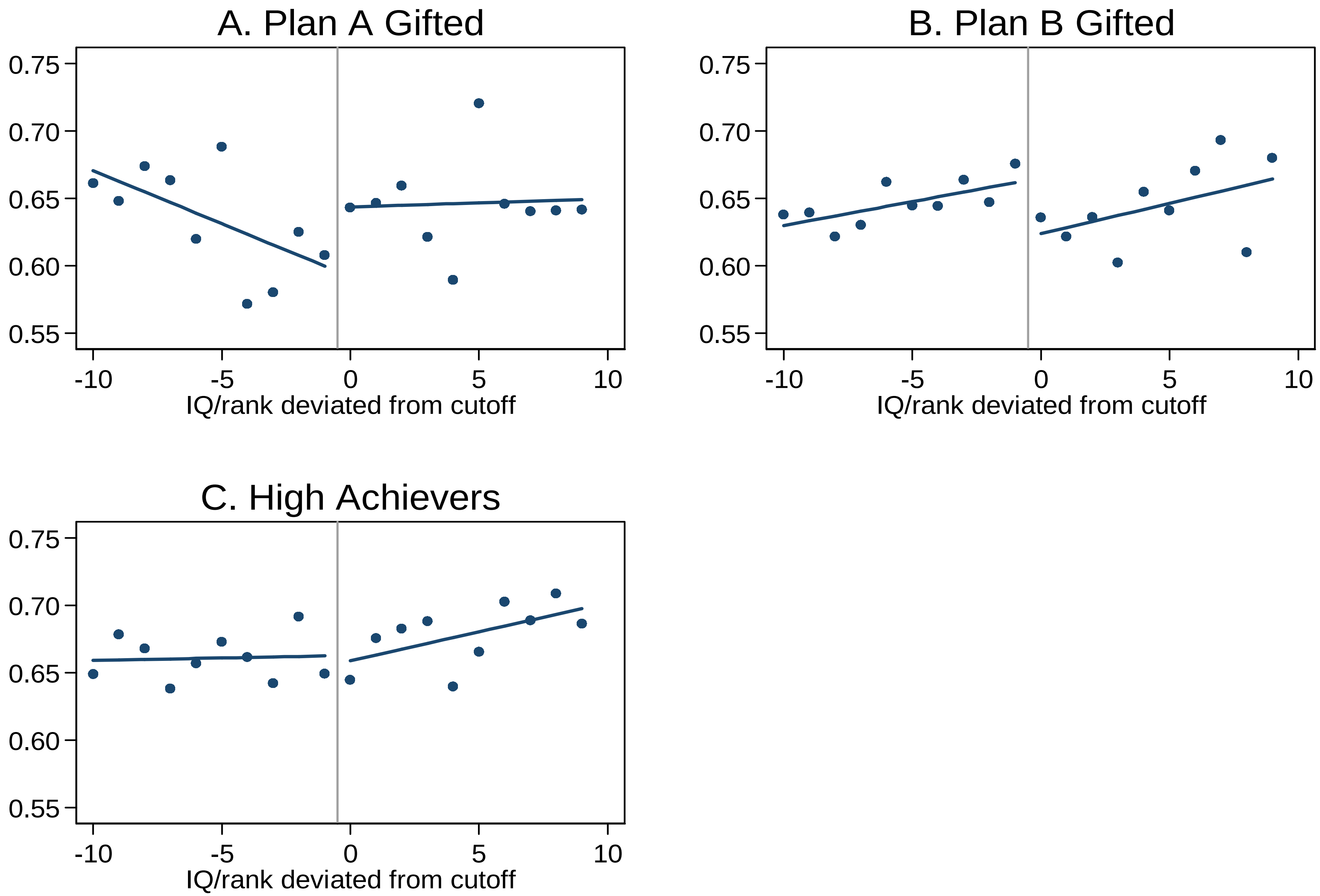
Appendix Figure 1. Fraction of IQ-tested students/potential high achievers for whom 4th grade outcomes are observed



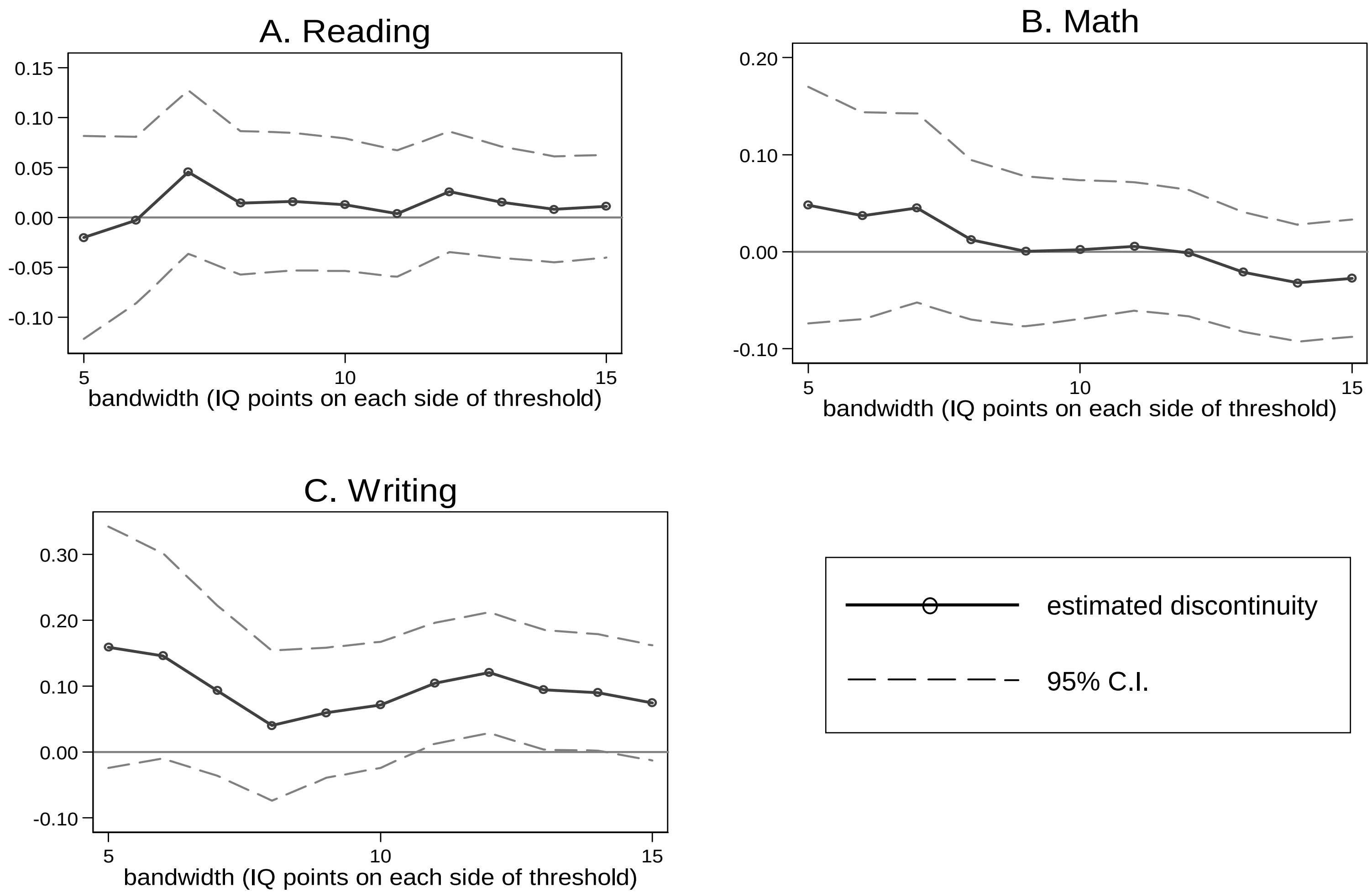
Appendix Figure 2. Estimated discontinuities in fourth grade scores from local linear regressions with varying bandwidths, Plan A sample



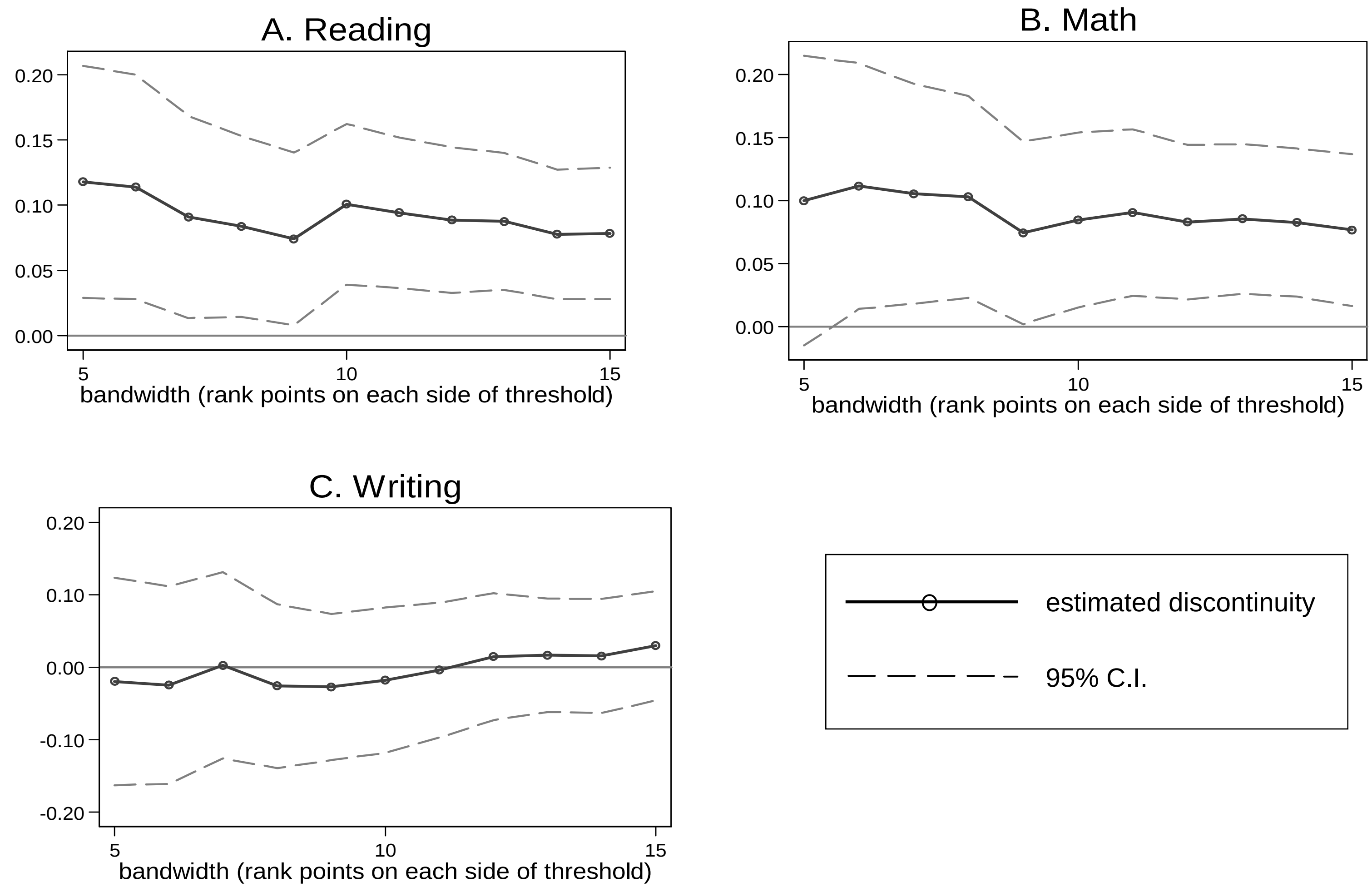
Appendix Figure 3: Index of self-reported satisfaction based on fourth grade survey responses



Appendix Figure 4. Estimated discontinuities in fourth grade scores from local linear regressions with varying bandwidths, Plan B sample



Appendix Figure 5. Estimated discontinuities in fourth grade scores from local linear regressions with varying bandwidths, High achiever sample



Appendix Figure 6: Discontinuities at Rank Threshold for Admission to 4th Grade Gifted Class for Subsample of High Achievers Observed in 5th Grade