

NBER WORKING PAPER SERIES

CAN UNIVERSAL SCREENING INCREASE THE REPRESENTATION OF LOW
INCOME AND MINORITY STUDENTS IN GIFTED EDUCATION?

David Card
Laura Giuliano

Working Paper 21519
<http://www.nber.org/papers/w21519>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2015

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 Attila Lindner, Carl Nadler and Sydnee Caldwell for their expert assistance. Special thanks to Hedvig Horvath for her input at many stages of this project. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D110019 to the National Bureau of Economic Research. The opinions expressed are those of the authors and do not represent views of the NBER or the U.S. Department of Education.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by David Card and Laura Giuliano. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Can Universal Screening Increase the Representation of Low Income and Minority Students in Gifted Education?

David Card and Laura Giuliano

NBER Working Paper No. 21519

September 2015

JEL No. I21

ABSTRACT

Low income and minority students are under-represented in gifted education programs. One explanation for this pattern is that the usual process for identifying gifted students, through parent and teacher referrals, systematically misses many potentially qualified disadvantaged students. We use the experiences in a large urban school district following the introduction of a universal screening program for second grade students to study this hypothesis. With no change in the standards for gifted eligibility the screening program led to large increases in the fractions of economically disadvantaged students and minorities placed in gifted programs. Comparisons of the newly identified gifted students with those who would have been placed in the absence of screening show that blacks and Hispanics, free/reduced price lunch participants, English language learners, and girls are all systematically "under-referred" in the traditional parent/teacher referral system.

David Card

Department of Economics

549 Evans Hall, #3880

University of California, Berkeley

Berkeley, CA 94720-3880

and NBER

card@econ.berkeley.edu

Laura Giuliano

Department of Economics

University of Miami

P.O. Box 248126

Coral Gables, FL 33124-6550

lgiuliano@miami.edu

Low income and minority students are substantially under-represented in gifted and talented education programs in the U.S. (see e.g., Ford, 1998; Donovan and Cross, 2002).¹ While some of the gap may be due to differences in the cognitive development of students from different backgrounds, there is also evidence that the traditional referral-based system for identifying gifted children tends to overlook potentially eligible students from disadvantaged families (e.g., Woods and Achey, 1990; Figlio, 2005; Ehloweris et al., 2005). This raises the question of whether a comprehensive screening program might be able to identify significant numbers of poor and minority students who qualify for gifted programs but would go undetected in a referral-based system.

We answer this question using data from a unique natural experiment conducted by one of the largest and most diverse school districts in the country (hereafter "the District"). State laws governing the District's policies dictate that non-disadvantaged students must achieve a minimum of 130 points on a standard IQ test to be eligible for gifted programs. English language learners (ELL's) and free-or-reduced price lunch (FRL) participants are subject to a lower 116 point threshold, known as "Plan B" eligibility. Even with this lower bar, however, the District's gifted student population in the early 2000s was largely comprised of white children from higher-income neighborhoods. Only 28% of gifted students in third grade were black or Hispanic, compared to 60% of all students in the District. Thirteen of the regular elementary schools in the District had no gifted children in third grade in 2004 or 2005, while the gifted rate was nearly 10% at the 13 schools with the lowest rates of FRL participation.

In response to these disparities the District introduced a universal screening program in spring 2005. Under this program, all second graders completed a non-verbal

¹ For example, in 2006 the national average participation rates of K-12 students in gifted and talented programs was 8.0% for whites, 3.8% for blacks, and 4.2% for Hispanics (U.S. Department of Education 2010, Table 49). In the large urban district we study, the gifted program participation rate for students who receive free or reduced price lunches was 1.9% in the 2004-5 period, while the rate for other students was 5.5%.

ability test, and those scoring above a threshold of 130 points (for non-disadvantaged students) or 115 points (for ELL and FRL participants) were eligible for referral to a District psychologist for IQ testing.² As in earlier years students could still be nominated for testing by parents or teachers, and families could have their children tested privately and submit the IQ scores to satisfy the state mandate. The screening program was therefore intended to supplement the traditional referral system, helping to narrow the gaps in referral rates between students from different backgrounds. Importantly, the IQ thresholds and other requirements for gifted eligibility were unchanged, so any increase in the number of students identified as gifted following the introduction of the program can be attributed to the screening effort, and not to a relaxation of the standards for gifted status.

Comparisons across cohorts of third-graders, illustrated in Figure 1, suggest that the introduction of universal screening led to large increase in the number of gifted students in third grade in the District. In contrast, the gifted rate in a matched comparison group of schools from other districts in the same state was quite stable.³ As a result of financial pressures caused by the Great Recession the District cut its funding for IQ testing in 2007, and suspended the universal screening program altogether in 2010. Thus, the original program was short-lived, and by 2011 the fraction of gifted third graders in the District was back down to the level of 2004-2005. At the same time the gifted rate in the comparison schools began to increase after 2007, reflecting program changes in other districts.

In light of the history of the District's screening program and the striking patterns in Figure 1, in this paper we focus on simple "pre/post" comparisons between third graders in 2004-2005 (the two cohorts immediately before the introduction of universal screening) and those in 2006-2007 (the two cohorts immediately after). We present

² As discussed below, the District used the Naglieri Non-verbal Ability Test (NNAT), rescaled to have mean 100 and standard deviation 15, for screening.

³ Figure 1 is based on publicly available school-level data for schools in the District and in other districts in the same state. Further information is presented later in the paper.

evidence confirming that the 2004-2005 ("pre") cohorts form a credible comparison group for the 2006-2007 ("post") cohorts. We use differences between these groups to measure the impact of the program on gifted participation rates and to characterize two key groups of participants: students who were classified as gifted in the post cohort and would have been identified in the pre cohort; and those who were classified as gifted in the post cohort but would have been overlooked in the pre cohort. These are equivalent to the "always takers" and the "compliers" in a standard analysis of experimental designs with incomplete compliance (e.g., Imbens and Angrist, 1995). By studying the characteristics of the compliers and their distribution across schools we gain important new insights into the types of students who would normally "fall through the cracks" of the traditional referral system for gifted education programs.

Our analysis points to four main conclusions. First, as shown in Figure 1, the introduction of the screening program led to a large increase in the fraction of students classified as gifted, with no cut in the standards for gifted certification. Second, the newly identified gifted students were disproportionately poor, black, and Hispanic, and more likely to have parents who speak a language other than English. They were also concentrated at schools with high shares of poor and minority students and low numbers of gifted students prior to the program. Thus, the experiences of the District confirm that universal screening can significantly increase the representation of disadvantaged students in gifted and talented programs. Third, the distribution of IQ scores for the newly identified students was very similar to the distribution among those who were identified under the referral system, particularly among students who qualified under the Plan B eligibility standard. The compliers included many students with IQ's significantly above the minimum eligibility threshold, implying that even relatively high ability students from disadvantaged backgrounds were being overlooked under the traditional referral system. Finally, comparisons of the achievement gains made between the end of third grade and the end of fourth grade suggest that the newly identified gifted students benefitted at least as much from participating in the

gifted program as the always taker group who would have identified under a referral system.

I. Background: The Referral Process and the Screening Program

Prior to the introduction of the screening program candidates for gifted status in the District were identified through an informal process of parent and teacher referrals. Referrals were concentrated in first and second grades to ensure that qualified students could begin receiving gifted services by third grade.⁴ Nominees were placed in a queue for a free evaluation and IQ test by a District psychologist. Parents could also bypass the queue by hiring a private psychologist and submitting the IQ score directly. Students with IQ's above the relevant threshold were eligible for gifted status,⁵ with the final determination based on parent and teacher inputs and scores on a checklist verifying that the student showed evidence of "gifted indicators," including motivation, creativity, and adaptability.⁶ The referral and evaluation process remained unchanged after the introduction of the screening program and is still in place today.

There is a thriving private market for IO testing in the District, with many psychologists offering first-time IO tests and re-testing for children who failed to meet the state mandated standards on an earlier test.⁷ Advertisements posted by these psychologists suggest that their target market is mainly non-disadvantaged families whose children have to meet the 130 point IQ threshold (known as the "Plan A" eligibility rule). The importance of private testing for these families is underscored by an analysis of the testing records for students who were placed in the gifted program by

⁴ More information on the District's gifted program is presented in Card and Giuliano (2014). In most schools gifted programming starts in third grade, and in fourth and fifth grades students are placed in separate classes designated for gifted and high-achieving students.

⁵ The state law actually allows students to score within a standard error (3 points) of the IQ threshold and still be placed in the gifted program. A small fraction of gifted students have scores 1.2 or 3 points below the mandated threshold.

⁶ The eligibility checklist for Plan B students also includes measures of academic achievement and family background factors.

⁷ A typical price for an evaluation and IQ test was \$700-\$1000 in 2013.

end of third grade in the 2004 and 2005. Among the Plan A eligible students, 20% were qualified on the basis of a private IQ test. In contrast, less than 1% of the Plan B eligible students were qualified through a private test.

Under the screening program, introduced in spring 2005, all second graders in the District completed the Naglieri Non-Verbal Ability Test (NNAT) (see Naglieri and Ford, 2003). The NNAT test -- which is similar to Raven's Progressive Matrix Test -- takes less than an hour to complete and was administered by second grade teachers during regular class time. Non-disadvantaged students scoring above 130 points on the test, and ELL/FRL students scoring above 115 points were eligible for referral for IQ testing.⁸ The aim of the program was to identify high-ability students and have them tested relatively quickly so they could begin receiving gifted services in third grade. Consistent with this objective, our analysis of records for gifted students in the third grade cohorts of 2006 and 2007 suggests that most completed an IQ test by December of the preceding year.

A key feature of the screening program was the need to increase the number of IQ tests performed by District psychologists. We estimate that about 1,300 additional IQ tests were conducted by District staff in the summer and fall after the first screening test in spring 2005, and a similar number of extra tests were performed in 2006. Since each test takes approximately 3 hours, the cost in overtime payments for testing staff was relatively large. In response to a budget crisis in 2007, the District decided to cut the overtime budget for testing staff. The result, clearly evident in Figure 1, was a sharp decline in the fraction of students placed in the gifted program by the end of third grade. Continuing budget pressures led the District to suspend the screening program altogether in 2011.

⁸ Teachers were supposed to fill out a checklist of qualifications and only refer students with sufficiently high scores on the checklist, in anticipation of the checklist requirement for gifted placement.

II. Data Description and Research Design

Our main data source for evaluating the effects of the District's screening program consists of longitudinal records for elementary students who were enrolled in third grade during spring semester of years from 2004 to 2007. We restrict attention to students who attended one of the 140 larger elementary schools in the District, excluding those at charter schools and other special schools. For our analysis sample we also exclude students who repeated third grade. Thus every student was in third grade during the 4-year sample period and would have written the screening test (if one was available) in the spring of the previous calendar year.

Column 1 of Table 1 presents some descriptive information on the roughly 40,000 third graders in our analysis sample during the pre-screening period (2004 and 2005).⁹ The District's student body is racially diverse, with 35% white non-Hispanic students, 34% black non-Hispanics, 25% Hispanics, and 3% Asians. Some 45% of third graders in these cohorts were eligible for free or reduced price lunches (FRL), and 11% were English language learners (ELL). Altogether 49% of third graders were either FRL or FRL and were therefore eligible for gifted status under the state's Plan B standard. The other 51% had to meet the regular (Plan A) eligibility requirements.

Columns 3 and 5 show the characteristics of the Plan A and Plan B eligible students, respectively. White and Asian students are over-represented in the Plan A group while blacks and Hispanics are over-represented in the Plan B group. We also report the overall fractions of FRL and black or Hispanic students at the schools attended by the two groups. Plan B eligibles attended schools with higher fractions of minority and FRL participants, reflecting residential segregation patterns in the District and the fact that school assignment is neighborhood based.

The lower panel of the Table presents information on IQ testing rates, mean IQ scores (for students who are tested), and gifted status as of the end of third grade.

⁹ For convenience, from now on we refer to cohorts by the year that they were in spring semester of third grade.

Overall about 16% of third graders in the District in 2004-2005 had an IQ score on record, with an average score of just over 100 points. Not all these students were tested to screen for gifted status. Many students are tested for diagnostic purposes, including the detecting of learning disabilities. Overall about 4% of students in the 2004-2005 cohorts had an IQ score above their respective eligibility threshold for potential gifted status, and 3.3% were classified as gifted. Despite the lower threshold for Plan B eligible group the gifted rate was still much lower for these students (1.5%) than for the Plan A group (5.1%).

Columns 2, 4, and 6 report the characteristics of third grade students in the District in the two years after the introduction of the screening program. Overall, these students are very similar to those in the two earlier cohorts, with the only notable difference being a small decline in the fraction of whites, offset by small rises in the fractions of black and Hispanic students. Given this stability, and the relatively smooth trend in economic conditions in the District's metropolitan area between mid-2004 and mid-2007,¹⁰ we believe that student outcomes in 2004-2005 represent a credible counterfactual for the outcomes of students in 2006-2007, particularly if we adjust for a trend to capture long-run demographic shifts and the gradually improving economic climate.

The validity of this interrupted time series design is supported by the patterns in Figure 1, which show that between 2004 and 2007 the fraction of gifted children in third grade in a matched comparison group of schools from the same state was relatively stable (3.5% in 2004, 3.6% in 2005, and 3.7% in 2006 and 2007). Thus, for most of our analysis below we use a simple pre/post "differences" design, though we also report some results based on comparisons of publically available school/grade level data from other districts to verify that the fraction of gifted third graders in other medium and large districts were very stable between the pre and post periods.

¹⁰ The unemployment rate in the local MSA was 5.6% in March 2004, 3.7% in March 2005, 2.7% in March 2006, and 2.8% in March 2007.

III. Impacts of the Screening Program on Testing and Gifted Rates

a. Screening Test Outcomes, IQ Testing, and Gifted Placement

The screening tests in spring 2005 and spring 2006 revealed that approximately 13 percent of Plan A-eligible students and 18 percent of Plan B-eligible students had NNAT scores above the 130 and 115 point thresholds, respectively. Unfortunately, we do not have direct information on which students were referred for IQ testing as a result of their scores. As shown in columns 2, 4, and 6 of Table 1, however, the overall fraction of third graders with an IQ test rose by 8 percentage points (ppt.) between the pre and post cohorts, with a 5 ppt. gain for Plan A eligible students and an 11 ppt. gain for the Plan B eligibles. This extra testing resulted in a 3 ppt. rise in the overall fraction of third graders with an IQ test score above the threshold for gifted eligibility, with a 2 ppt. gain for the Plan A eligibles and a 4 ppt. gain for the Plan B group. Most importantly, the fraction of students placed in the gifted program rose by 2.2 ppt. relative to the pre cohorts, with a 1.5 ppt. gain for the Plan A group and a 2.9 ppt. gain for the Plan B group.¹¹

Further insights into the impacts of the screening program are revealed by Figures 2a and 2b, which show how IQ testing rates and gifted rates were related to NNAT screening test results for students in the post cohorts.¹² The design of the screening program implies that IQ testing rates and gifted rates should jump discontinuously at the screening thresholds if the program operated as intended. To aid in interpreting the figures, we also show local linear regression discontinuity style

¹¹ A complicating issue is that some Plan B-eligible students in second grade change status in third grade. These are mainly English learners who transition out of the language program after second grade. These students can be referred for IQ testing and placed in the gifted program as Plan B eligible, but recorded as Plan A eligible in third grade. As discussed below, we estimate that this group accounts for about one-third of the newly identified students coded as Plan A eligible in third grade.

¹² Note that the figures are based on data for students who were in the District in second grade at the time of the screening test and enrolled in third grade in a District school the next year.

models fit separately to observations above and below the respective screening thresholds. Estimates of these models are reported in Appendix Table 1.

Looking first at the results in Figure 2a for Plan A eligible students, students who scored just under the screening test threshold of 130 points had a 38% probability of being IQ tested, whereas those who scored just above the threshold had a 58% chance of being tested. Thus, passing the screening test threshold was associated with a 20 percentage point increase in the probability of IQ testing for Plan A eligible students. Nevertheless, even non-disadvantaged students with NNAT scores well above the threshold had a less than 100% chance of being IQ tested within the next year.

The corresponding jump in the probability of having an IQ score above the gifted program minimum is 7 ppt., implying that about one-third of Plan A eligible students who were referred for IQ testing as a result of barely passing the screening threshold had an IQ score of 130 points or higher. There is also about a 7 ppt. jump in the probability of being placed as a gifted student, which suggests that most of the newly identified Plan A eligible students with NNAT scores just above the screening threshold were actually placed in the gifted program.

The results in Figure 2b for Plan B eligible students show some important differences with the results for the Plan A group. Just to the left of the 115 point threshold the IQ testing rate for the Plan B group was about 10% - only 1/4 of the testing rate for Plan A eligibles with scores just under 130 points.¹³ This disparity is consistent with concerns that many disadvantaged students with ability levels around the threshold for gifted status were being overlooked in the traditional referral system. In contrast, the IQ testing rate for Plan B eligibles with screening scores just above the 115 point threshold is about the same as the rate for Plan A eligibles just above their threshold (around 60% in both cases). This suggests that the "marginally qualified" Plan

¹³ In fact it seems likely that most of the Plan B group just below the screening test threshold who had an IQ test on record were tested for some reason other than to check for gifted eligibility, so the disparity is even larger.

A and Plan B groups were treated equally in the post-screening referral process, effectively closing the gap in referral rates for students just above the screening bar.

Looking at the fraction of the Plan B group with IQ scores of 116 or higher, and the fraction who were placed in the gifted program, both are very close to zero for students with an NNAT score just under the Plan B screening threshold (115 points). For those with a screening score just above the threshold, on the other hand, the fraction with an IQ score of 116 or more is 16% and the fraction classified as gifted is 10%. Thus, about one quarter of the newly identified Plan B group who were IQ tested as a result of an NNAT score just above the threshold had IQ's of 116 points or higher -- a "success rate" somewhat below the 30% rate for Plan A eligibles with NNAT scores just above 130. Only 60% of the marginal Plan B group with IQ scores above 116 were actually placed in the gifted program, however, compared with a nearly 100% placement rate for marginal Plan A students with IQ scores above 130.

To help better understand the determinants of which disadvantaged students were actually IQ tested as a result of the screening program, and which of those with high enough IQ's were placed in the gifted program, we fit a series of descriptive models to the universe of Plan B students in the post cohorts who had NNAT scores above the 115 point screening threshold. We fit separate models for being IQ tested; having an IQ score of 116 or higher if tested; and being placed in the gifted program if IQ-qualified. The results are reported in Table 2. For each of the three outcomes we present three specifications: a basic model with only demographic characteristics of students and their NNAT scores; a second model with school dummies that control for potential unobserved factors across different schools (e.g., teacher or principal enthusiasm for the screening program); and a third model with school dummies and a control for the student's third grade scores on statewide achievement tests.¹⁴

¹⁴ Statewide testing only starts in third grade so we have no second grade achievement scores. Third grade scores may be affected by gifted placement, though other research (e.g., Card and Giuliano, 2014; Bui, Craig and Imberman, 2011) suggests the effect is probably very small.

The models for IQ testing (in columns 1-3) show that students with higher NNAT's were more likely to be tested -- consistent with the simple patterns in Figure 2b. Testing rates were also higher for Asian students, younger students, and those with higher third grade achievement scores. Models for the likelihood of having an IQ score above the state threshold (columns 4-6) suggest that this probability was higher for students with higher NNAT scores, lower for older students, black students and female students, and much higher for those with higher achievement levels. Finally, models for the probability of placement in the gifted program conditional on having an IQ score above the minimum threshold (columns 7-9) suggest that these rates were slightly higher for black and Hispanic students, higher for Asian students, lower for ELL students, and much higher for students with higher levels of achievement.

To illustrate the implications of the achievement score effects from the models in columns 3, 6 and 9 of Table 2, consider the effect of a 1 standard deviation higher achievement score (holding constant NNAT scores and other factors). This would lead to a 7 ppt. increase in the likelihood of being IQ tested, a 22 ppt. increase in the likelihood of actually scoring above the 116 point threshold on an IQ test, and a 21 ppt. increase in the likelihood of being placed in the gifted program, conditional on scoring at least 116 on the IQ test. These estimates suggest that the referral, testing, and placement system operated by the District after the introduction of the screening test selectively pushed higher-achieving Plan B eligibles into the gifted program, while holding back "under-achievers" with adequate NNAT scores or IQ scores but lower achievement levels. This pattern is broadly consistent with the guidelines teachers were expected to use in deciding which students to refer for IQ testing once the screening test scores were available, and with the checklist used in combination with IQ scores to determine gifted placement of Plan B students.

b. Net Impacts of the Screening Program on Rates of Gifted Placement

With this background we now turn to estimates of the impact of the screening program on the fraction of gifted students in the District, summarized in Table 3. We begin in Panel A with estimates based on the same publicly available school/grade data used in Figure 1.¹⁵ Row 1 shows the mean fraction of third grade gifted students in the District in the pre and post years, and the change in this fraction, which is 1.7 ppt.. Row 2 shows the difference in the fraction gifted between schools in the District and a comparison group of schools in other districts in the state.¹⁶ This gap corresponds to the difference in the fraction gifted between District schools and the comparison group schools in Figure 1. The difference of differences between District schools and comparison group schools is 1.6 ppt., with a standard error (clustered by school) of 0.3 ppt.. We also fit a model which allows gifted rates at schools in the state to have a trend over time. The difference in differences from this model, presented in column 4, is virtually the same as the entry in column 3.

A possible concern with the estimates in rows 1 and 2 is that their precision is over-stated by the presence of unobserved district-wide factors that cause the fraction of gifted children to vary from year to year. Such factors are not incorporated in the standard errors reported in the table. To assess the likely importance of district-level error components in the fraction of gifted students, we constructed the empirical distribution of changes in the average fraction of third graders in gifted programs from 2004-5 to 2006-7 for the 23 larger districts in the state (those with at least 20

¹⁵ These data are available from the state testing agency, and report the numbers and characteristics of students who took the statewide achievement tests in each school, grade, and year. For consistency with our student level data, we select larger elementary schools (at least 40 third graders in each year) that are not classified as charter schools.

¹⁶ We fit a propensity score model to the pooled sample of larger District schools and larger non-charter schools in other districts, and used the estimated scores to weight the data for the comparison schools. The variables included in the propensity score model were: mean fraction gifted in 2004-5, mean FRL participation rate of third graders in 2004-5, 2006-8, and 2009-11, mean fraction of blacks and Hispanics among third graders in 2004-5, 2006-8, and 2009-11, and mean number of third grade students in all years 2004-2011.

elementary schools), including the District (which, with 140 elementary schools is one of the largest in the state).¹⁷ We show this distribution in Figure 3, along with the confidence intervals associated with the changes in the fraction of gifted students in each district. Two features are clear from the figure. First, among other larger districts the variability in the fraction gifted between the pre and post cohorts is relatively small. Second, relative to this distribution the change in the fraction gifted in the District is a clear outlier.¹⁸ We conclude that a simple within-District differences analysis provides a plausible estimate of the effect of the screening program on the overall fraction of gifted third graders.

Panel B of Table 3 shows within-District differences estimates based on student level data for schools in the District. For each student group we present a simple difference estimate in column 3 and a trend-adjusted difference in column 4. The results for all third grade students (row 3) show a slightly larger increase than the aggregated school-level data, but the trend adjusted difference is nearly identical to our difference-in-differences estimate. In the lower rows we present estimated impacts for various subgroups of third graders. The trend-adjusted impacts (which we prefer) are relatively small for Plan A eligibles and whites, while the impacts for Plan B eligibles, blacks, and Hispanics are all relatively large.¹⁹

To interpret these estimated impacts it is helpful to translate the percentage point effects into numbers of gifted third graders. The trend adjusted difference estimates suggest that the screening program added about 300 gifted students to the total in the District, consisting of roughly 60 (20%) Plan A eligibles and 240 (80%) Plan B

¹⁷ We exclude one other district which adopted a screening program in 2005. The change in the gifted fraction in this district is 1.76 ppt., just a little larger than the change in the District.

¹⁸ The empirical mean and standard deviation of the changes in fraction gifted across the 22 other districts in the state are -0.001 and 0.007, respectively. The observed change in the District is 2.65 standard deviations from the mean for the other districts, which can be interpreted as a t-test statistic.

¹⁹ As shown in Appendix Figures 1a and 1b, there are clearly positive trends in the gifted rates of various subgroups of student in the District.

eligibles. The added Plan B's represent a 180% increase in the total number of disadvantaged gifted students in third grade relative to the numbers in 2004/2005. The percentage impacts on the numbers of black and Hispanic students in the District gifted program are also very large: around an 80% increase in the number of black participants, and a 130% increase in the number of Hispanic participants.

IV. A Closer Look at the Newly Identified Gifted Students

a. Student Characteristics

The evidence in Table 3 suggests that the changes in gifted participation rates caused by the District's screening program were very different across economic and racial groups. Table 4 presents a more systematic analysis that allows us to compare the newly identified gifted students -- the "compliers" whose gifted status was changed by the screening program -- to the students who would have been identified under the traditional referral system. We begin in columns 1 and 2 by showing the fractions of all gifted students in the pre cohorts and the post cohorts with the characteristic identified in the row heading. In the absence of the screening program or any trend factors affecting the total fraction of District students with the characteristic, the fractions would expected to be the same in the two cohorts. The characteristics of the complier group, in column 3, can be identified by comparing fractions in the two cohorts and adjusting for any trend across the 4 years in our sample.²⁰ Finally, in column 4 we show the difference in means between the compliers and the group of gifted students identified in the pre cohorts through the traditional referral system.

Several interesting patterns emerge from this analysis. As noted in the discussion of Table 3, about 80% of the compliers were Plan B eligibles. They are

²⁰ Specifically the estimates in columns (3) and (7) are calculated as the trend-adjusted change in the population share that is gifted and has the characteristic in question divided by the change in the population share that is gifted. To calculate the standard errors for these estimates and those in columns (4) and (8) we estimate the changes simultaneously using Stata's suest command and allow for clustering by school.

disproportionately black and Hispanic, and substantially less likely to have English-speaking parents. The newly identified group also attended schools with relatively high fractions of FRL participants and non-white students. Girls are more likely to be identified through screening than boys, whereas under the traditional referral system girls were under-represented in the District's gifted program.

The bottom two rows of the table show mean third grade math and reading scores and mean IQ scores of the various groups of gifted students. The newly identified students have about 0.4 standard deviation units lower third grade test scores than the always takers, and IQ scores that are about 0.5 standard deviation units below the scores of the always takers. In interpreting the IQ gap, however, it is important to keep in mind that most of the complier group are Plan B eligible, and therefore subject to a lower IQ threshold than the Plan A group who comprise the majority of the always takers.

Given this fact it is useful to focus on comparisons among the Plan B eligible subpopulation only. We report a parallel analysis for this group in columns 5-8 of Table 4. Note that the characteristics of the Plan B compliers, reported in column 7, are quite similar to the characteristics of all compliers reported in column 3, reflecting the fact that 80% of all compliers are Plan B eligible. In contrast, the Plan B always takers (summarized in column 5) are different from the overall sample of always takers (in column 1). Overall the gaps between these two Plan B groups are modest, though the compliers are less likely to be white, and more likely to be Hispanic, ELL, and to have non English speaking parents than the group of Plan B eligibles who were identified using traditional referrals. These gaps suggest that language may be a particularly important barrier preventing the identification of qualified gifted children in a referral-based system.

Looking at the last two rows of the table for the Plan B group, the compliers have lower achievement scores than the always takers, but about the same mean IQ scores. On average, it appears that the traditional referral system tends to overlook

disadvantaged students with modest achievement levels, regardless of their cognitive abilities. Teachers and parents may simply not recognize the abilities of many students – particularly those from poor and immigrant family backgrounds.

Further insights into the cognitive abilities of the compliers are provided in Figure 4, which plots the distribution of IQ scores of always takers and compliers in the Plan A and Plan B gifted populations in 2006 and 2007. Looking first at the data for the Plan B group, note that the two distributions are very fairly similar, though the compliers are less likely to have scores very close to the cutoff. Twenty percent of the compliers have IQ scores of 130s points or higher (versus 25% of the always takers), suggesting that many high-ability disadvantaged students are at risk of being overlooked in a traditional parent/teacher referral system.

The distributions of scores for compliers and always takers in the Plan A group are also interesting. About 30% of the compliers have scores under 125 points, versus 8% of the always takers. These are students who were classified as Plan B eligible at the time they were tested, but by end of third grade are Plan A eligible (mainly because they transition out of English language learning status). Taking account of this group, about 85% of all the newly identified gifted students were drawn from the Plan B pool. As with the Plan B compliers, many of the newly identified Plan A gifted students have IQ scores well above the minimum threshold, suggesting that the traditional referral system also misses some high ability non-disadvantaged students.

□

b. School Distribution

A troubling feature of the District's gifted education program prior to the introduction of the screening program was the very unequal distribution of gifted students across schools. This situation is illustrated by the blue line in Figure 5, which shows a "q-q" plot of the cumulative distribution of gifted 3rd grade students in the 2004-2005 cohorts among the 140 larger elementary schools in the District against the cumulative distribution of all 3rd grade students. (The schools are ranked by their

fraction gifted, with the first school contributing the largest relative share of gifted students). Fifty percent of gifted students in the pre cohort were at schools that enrolled only 18% of the entire District third grade population, while one half of all students in the District were at schools that enrolled a total of only 16% of the gifted students.

The effect of the introduction of the screening program is illustrated by the green line in the figure, which shows the q-q plot for the 2006 and 2007 cohorts. This line is much closer to the 45 degree line, implying a more equal distribution of gifted students across schools. In particular, after the introduction of screening all schools had at least 1 gifted 3rd grade student in 2006 or 2007, whereas in the pre cohorts, 13 schools had no gifted students in 2004 or 2005. The third (red) line in the figure shows the distribution of the compliers across schools. This line is (mostly) below the 45 degree line, implying that relatively few compliers were identified at the schools with the high fraction of gifted students in the pre cohort, whereas relatively more were identified at schools with low gifted rates in the pre cohort.

□

V. Academic Performance of the Newly Identified Gifted Students

One of the potential concerns with a screening program is that the newly identified students will be "mismatched" and will not benefit as much from the gifted program as traditionally referred students. Such arguments are often made against other outreach efforts, including programs to expand minority representation in elite universities (e.g., Arcidiacono et al., 2012). The newly identified students may also disrupt the quality of gifted services for other students, undermining the value of the program (Golden, 2004). Given the large increase in the number of gifted students following the introduction of the District's screening program, and the high fraction of poor and minority students among the newly identified group, it is useful to examine the potential effects on academic performance of the District's gifted student body.

To develop some simple evidence on this question we use the distribution of test score gains in reading and math between third and fourth grade for gifted students in the pre and post cohorts.²¹ Most gifted students in the District have already received some limited services by the end of third grade, such as individualized instruction in math or reading. In fourth grade, however, all gifted students are placed in designated classes for gifted students and other high achievers (see Card and Giuliano, 2014) and also receive a full year of gifted services. Arguably, the test score gains between third and fourth grade can provide useful, albeit noisy information on the degree to which students are successfully adapting to gifted education.²²

We began by measuring a reference distribution of test score gains, using the entire population of gifted students in the 2004 and 2005 cohorts. We then calculated the distributions of test score gains for gifted students in the 2006 and 2007 cohorts, classifying these students into 3 groups, based on the relative fraction of newly identified gifted students in the 2006/2007 cohorts at their school. In the lowest tercile schools there were very few newly identified gifted students, and we would expect the distribution of test score gains for students in these schools to closely match the reference distribution. In the highest tercile schools, on the other hand, most of the gifted students were in the newly identified group. If these students were "mismatched" to gifted education we would expect their distribution of test score gains to be shifted left relative to the reference distribution. Finally, at the middle tercile schools there was a combination of always takers and compliers. If, for example, the compliers were mismatched and their presence harmed the quality of education for the always takers, then we would also expect the distribution of gains to lie to the left of the reference distribution for students at these schools.

²¹ We use each student's standardized scores in reading and math in statewide tests at the end of third and fourth grades to measure the change in scores or the "gain score".

²² Very few students' standardized scores are top coded so we are not concerned that the tests have no power to measure achievement gains for gifted children.

The results of our comparisons are presented in Figure 6. We show the distributions of gains in each tercile group, along with the reference distribution. For each group we also note the share of newly identified gifted students. As expected, at the schools with the lowest fraction of newly identified gifted students we see that the distribution of test score gains is virtually identical to the reference distribution. This is reassuring since it also means that differences in the statewide test instruments used in different years do not have much effect on measured gain scores. In the second tercile, where about 35% of gifted students in the 2006-2007 cohorts were newly identified by screening, the distribution of gains is shifted slightly to the right relative to the reference distribution, showing no indication of a disruptive effect on the always takers. At the third tercile schools, where about three quarters of the gifted students in 2006-2007 were newly identified by the screening program, the distribution is also shifted slightly to the right relative to the reference distribution.

While a full analysis of the effects of participating in gifted education for different groups of students is beyond the scope of this paper, we view the results in Figure 6 as suggestive that if anything, the newly identified students benefitted even more from participating in gifted education than did the group of always takers who were identified under a traditional referral system. Moreover, the patterns at the middle tercile schools suggest that any spillover effect on the always takers were small.

VI. Conclusions

A longstanding concern about gifted education programs is the underrepresentation of disadvantaged groups, including minorities, non-native English speakers, and children from poor families. In response, a substantial body of research has focused on alternative methods for assessing giftedness that are less reliant on standard IQ tests (e.g., Renzulli, 1978; Borland and Wright, 1994; Naglieri and Ford, 2003; VanTassel-Baska, Feng and Evans, 2007). Over the past 30 years states and school districts have introduced multiple criteria for determining giftedness (e.g., grids using a

combination of IQ and achievement) and also adopted different thresholds for disadvantaged groups (McClain and Pfeiffer, 2012). Nevertheless, the fractions of black and Hispanic students in gifted programs remain far below the fraction of whites and Asians, and disparities between socioeconomic groups persist.

An alternative and complementary explanation for the under-representation gap is that the traditional referral process for identifying gifted students tends to systematically miss many potentially qualified disadvantaged students. The experiences in the District following the introduction of universal screening for second graders strongly support this hypothesis. With **no change** in the minimum standards for gifted status the screening program led to a 180% increase in the gifted rate among all disadvantaged students, with a 130% increase for Hispanic students and an 80% increase for black students.

A comparison of the newly identified gifted students to those who would have been identified even without screening shows that black and Hispanic students, free/reduced price lunch participants, English language learners, and girls were all systematically "under-referred" to the gifted program. Newly identified gifted students were more likely to come from schools in poor neighborhoods with relatively few gifted students, leading to a substantial equalization in gifted participation rates across schools. On average the newly identified students also had IQ scores that were similar to those of the always takers in the same eligibility group, though they had lower standardized achievement scores. We hypothesize that parents and teachers often fail to recognize the potential of many poor and immigrant children with less than stellar achievement levels, accounting for their likelihood of being under-referred.

An important limitation of our analysis is that it pertains to only a single school district. While the student population in the District is highly diverse, and arguably representative of the student population in many other large urban districts, the gifted program offered by the District has a number of distinctive features, including a minimum IQ threshold with a lower bar for free/reduced price lunch participants and

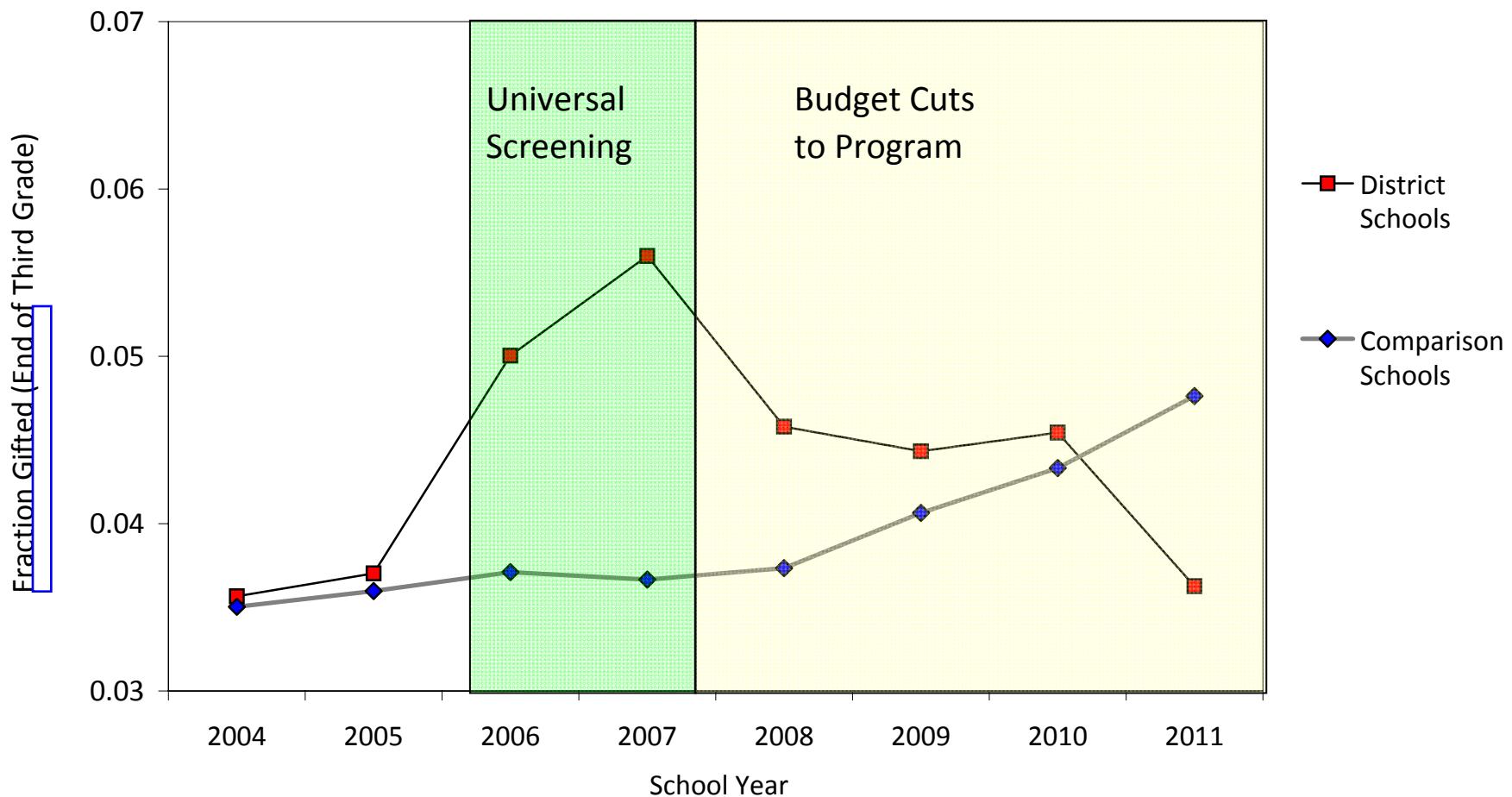
English language learners. Prior to the introduction of the screening program the District had relatively low gifted participation rates (around 3.5% for third graders), reflecting the importance of the IQ thresholds. At a minimum, however, our findings suggest that the under-representation of poor and minority students in gifted education is not due solely to the lower IQ scores of these students. A substantial share of the gap appears to be caused by the failure of the traditional parent/teacher referral system to identify high-ability disadvantaged students.

References

- Arcidiacono, Peter, Esteban Aucejo, Patrick Coate and V. Joseph Hotz (2012). "Affirmative Action and University Fit: Evidence from Proposition 209." NBER Working Paper 18523.
- Borland, James H. and Lisa Wright (1994). "Identifying Young, Potentially Gifted, Economically Disadvantaged Students." *Gifted Child Quarterly* 38(4): 164-171.
- Bui, Sa A., Steven G. Craig, and Scott A. Imberman (2011). "Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Achievement." NBER Working Paper 17089.
- Card, David and Laura Giuliano (2014). "Does Gifted Education Work? For Which Students?" NBER Working Paper 20453.
- Donovan, M. Suzanne and Christopher T. Cross (2002). *Minority Students in Special and Gifted Education*. Washington DC: National Academies Press.
- Elhoweris, Hala, Kagendo Mutua, Negmeldin Alsheikh, and Pauline Holloway (2005). "Effects of Children's Ethnicity on Teachers' Referral and Recommendation Decisions in Gifted and Talented Education." *Remedial and Special Education* 26(1): 25-31.
- Figlio, David N. (2005). "Names, Expectations, and the Black-White Test Score Gap." NBER Working Paper 11195.
- Ford, Donna Y. (1998). "The Underrepresentation of Minority Students in Gifted Education: Problems and Promises in Recruitment and Retention." *Journal of Special Education* 32(1): 4-14.
- Golden, Daniel (2004). "Boosting Minorities in Gifted Programs Poses Dilemmas." *Wall Street Journal*. April 7, 2004.
- Imbens, Guido W. and Joshua D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467-475.
- McClain, Mary-Catherine and Steven Pfeiffer (2012). "Identification of Gifted Students in the United States Today: A Look at State Definitions, Policies and Practices." *Journal of Applied School Psychology* 28: 59-88.

- Naglieri, Jack A. and Donna Y. Ford (2003). "Assessing Underrepresentation of Gifted Minority Children Using the Naglieri Nonverbal Ability Test (NNAT)". *Gifted Child Quarterly* 47(2): 155-160.
- Renzulli, Joseph S. (1978). "What Makes Giftedness? Re-examining a Definition." *Phi Delta Kappan* 60(3): 180-184.
- United States Department of Education, National Center for Education Statistics (2010). *Digest of Education Statistics*. Washington DC: GPO.
- VanTassel-Baska, Joyce, Annie Xuemie Feng and Brandy Evans (2007). "Patterns of Identification and Performance Among Gifted Students Identified Through Performance Tasks: A Three Year Analysis." *Gifted Child Quarterly* 51(3): 218-231.
- Woods, Sadie Bryant and Virginia H. Achey (1990). "Successful Identification of Gifted Racial/Ethnic Group Students Without Changing Classification Requirements." *Roeper Review* 13 (1): 1-26.

**Figure 1: Trends in Fraction Gifted by End of Third Grade
District Schools versus Matched Comparison Schools**



Note: school level data weighted by mean school enrollment in all years. Comparison schools weighted by propensity score \times enrollment. Screening program was eliminated in 2010, affecting students in 2011.

**Figure 2. Rates of IQ Testing and Gifted Placement
within One Year of Universal Screening**

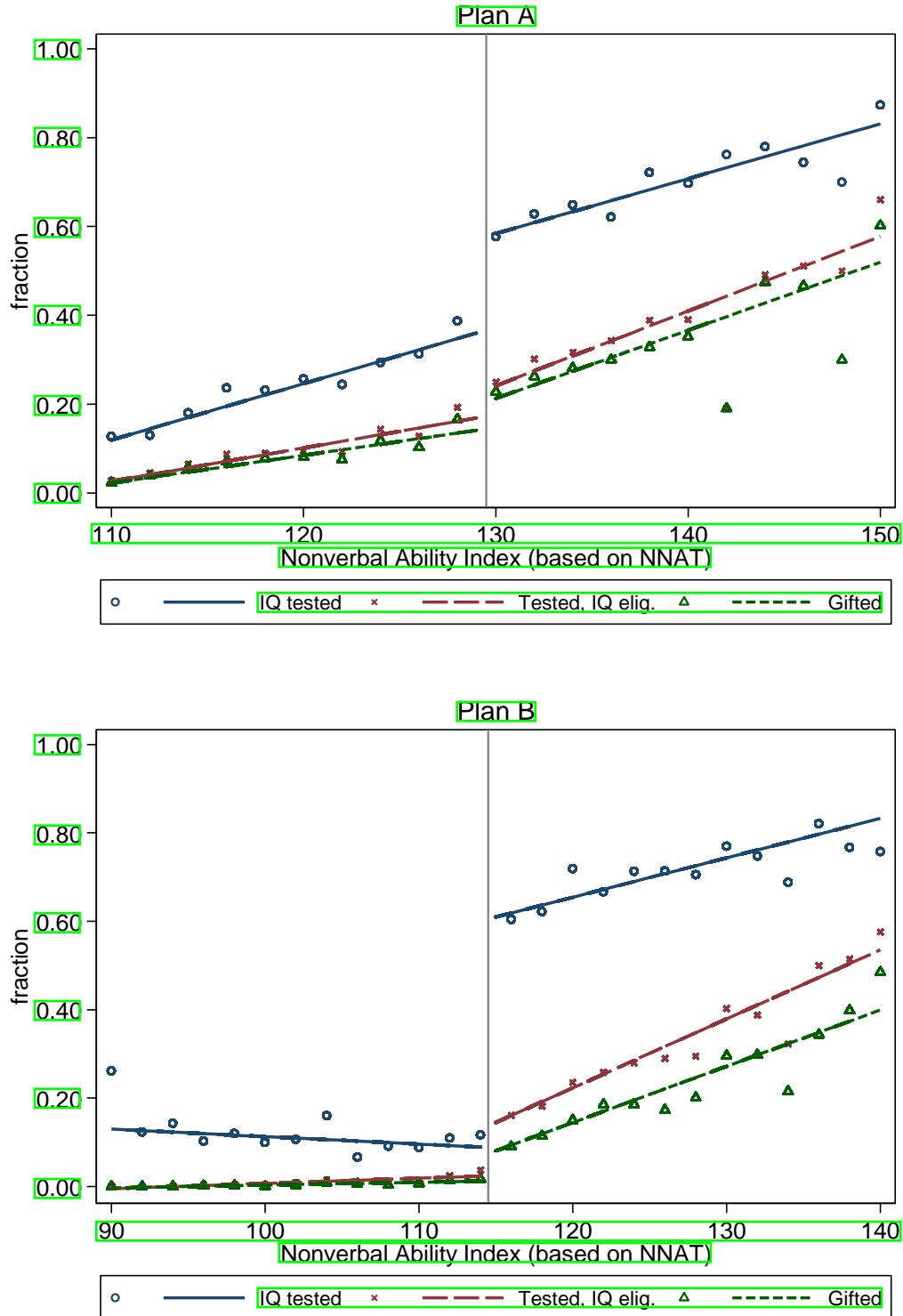
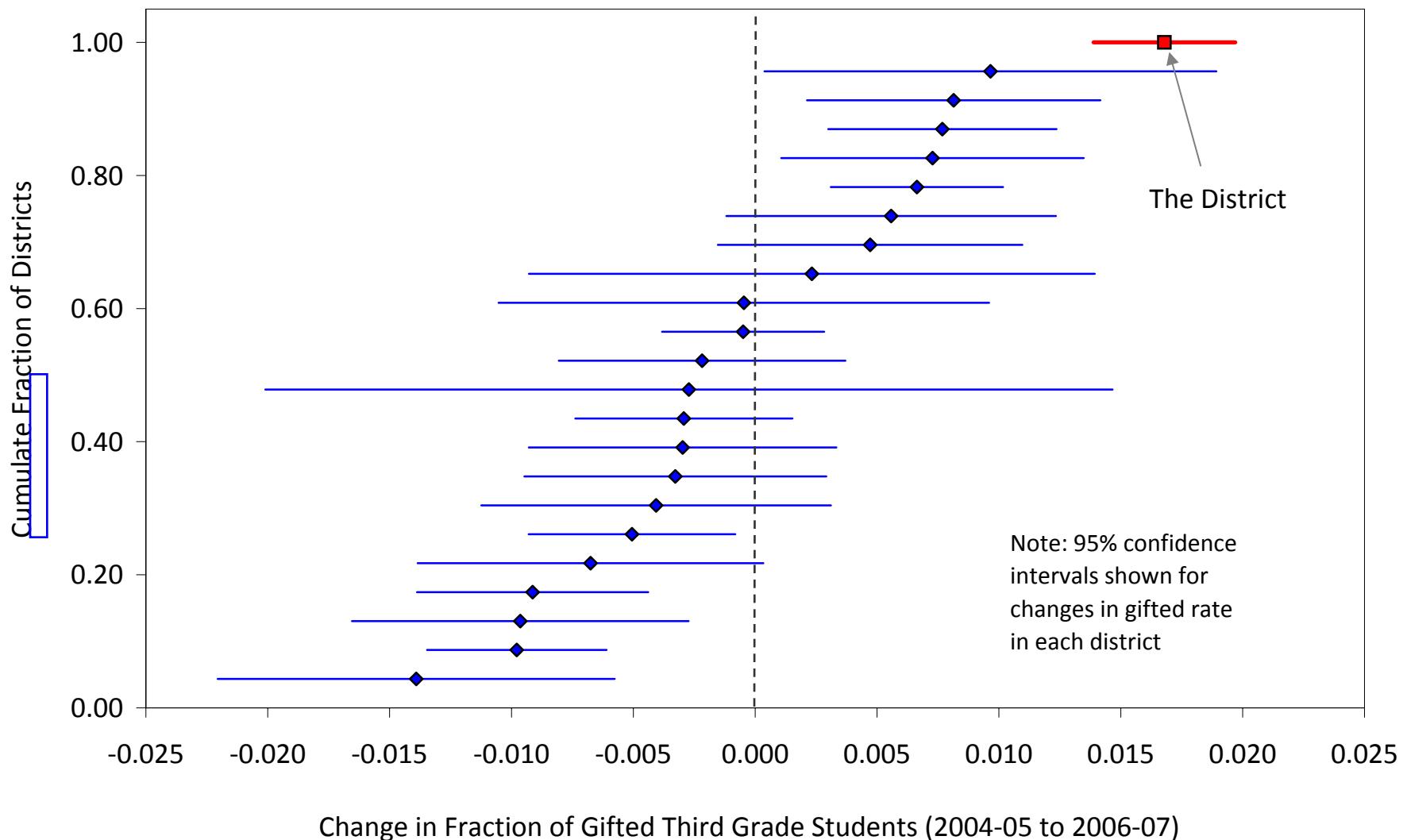


Figure 3: Cumulative Distribution of Changes in the Fraction of Gifted Third Grade Students
Between 2004-05 and 2006-07 (Districts with 20+ Elementary Schools)



**Figure 4: IQ Distributions of Always Takers and Students
Identified By Screening**

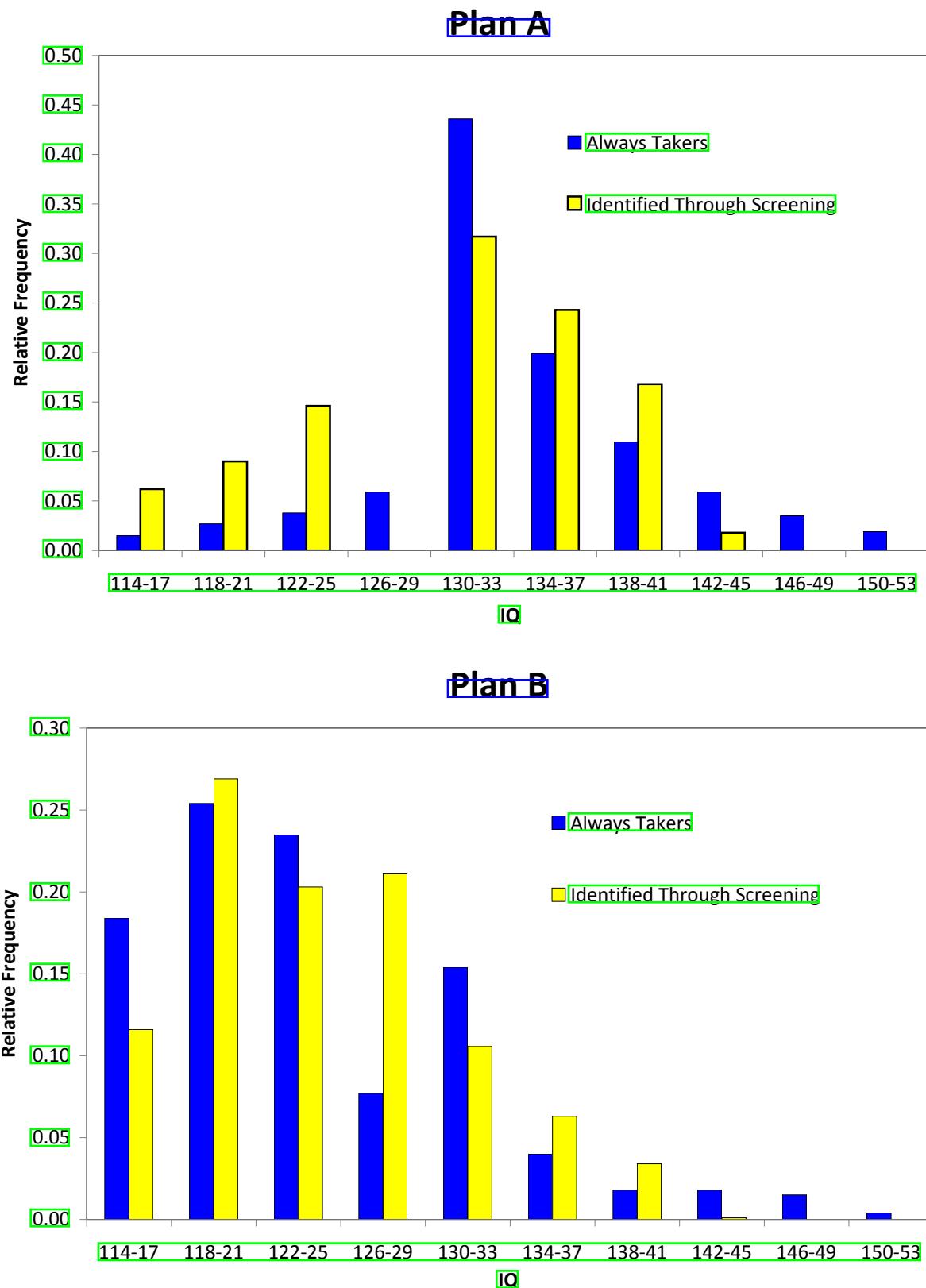


Figure 5: Distributions of Gifted Students Across Schools

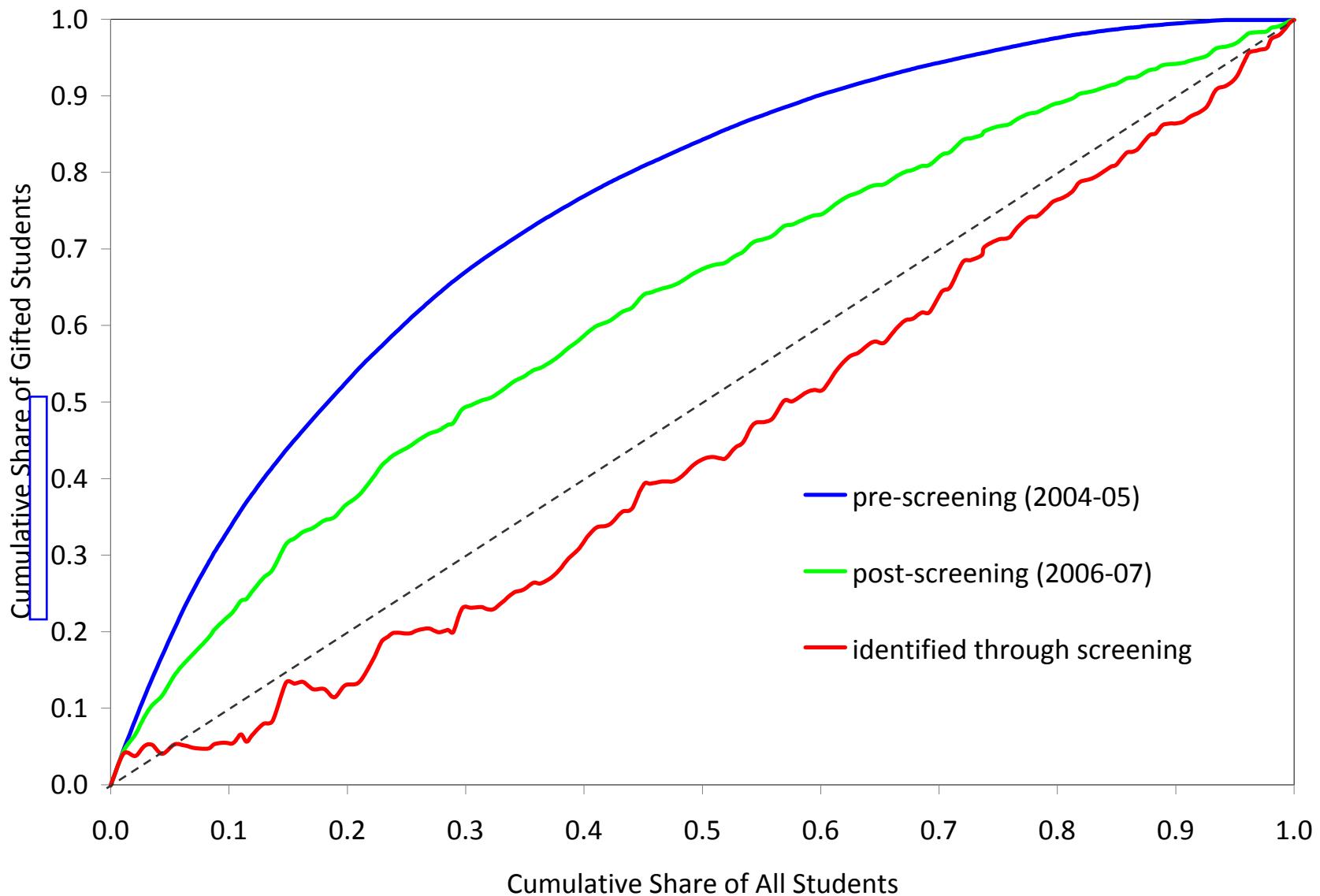


Figure 6: Distribution of Gifted Students' Test Score Gains, Pre-screening vs. Post-screening, by Tercile of School's Gifted Share Identified through Screening

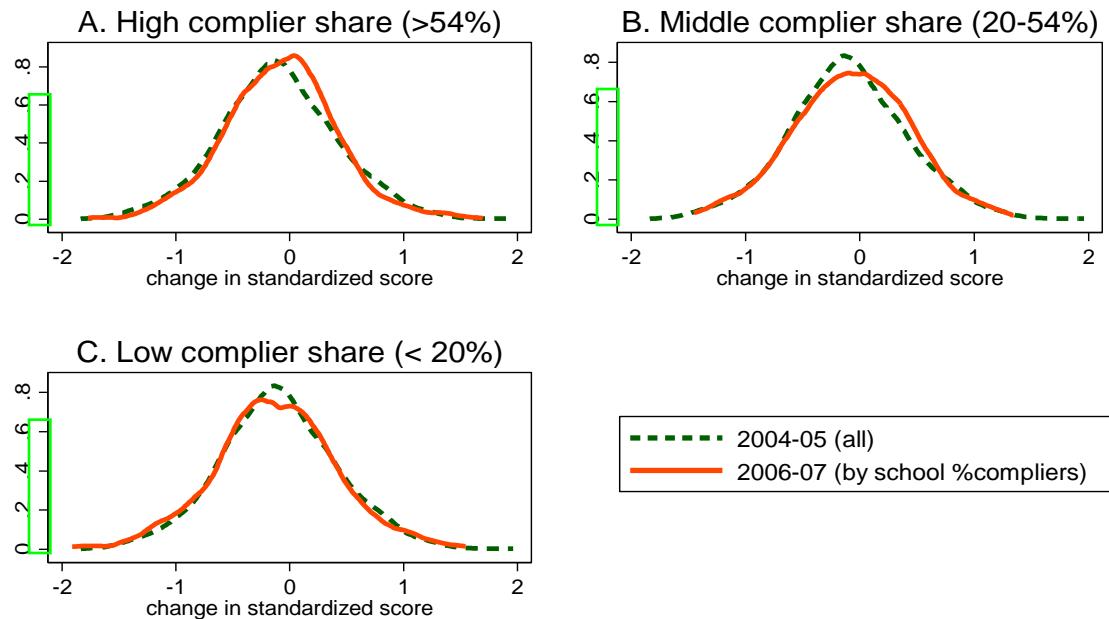
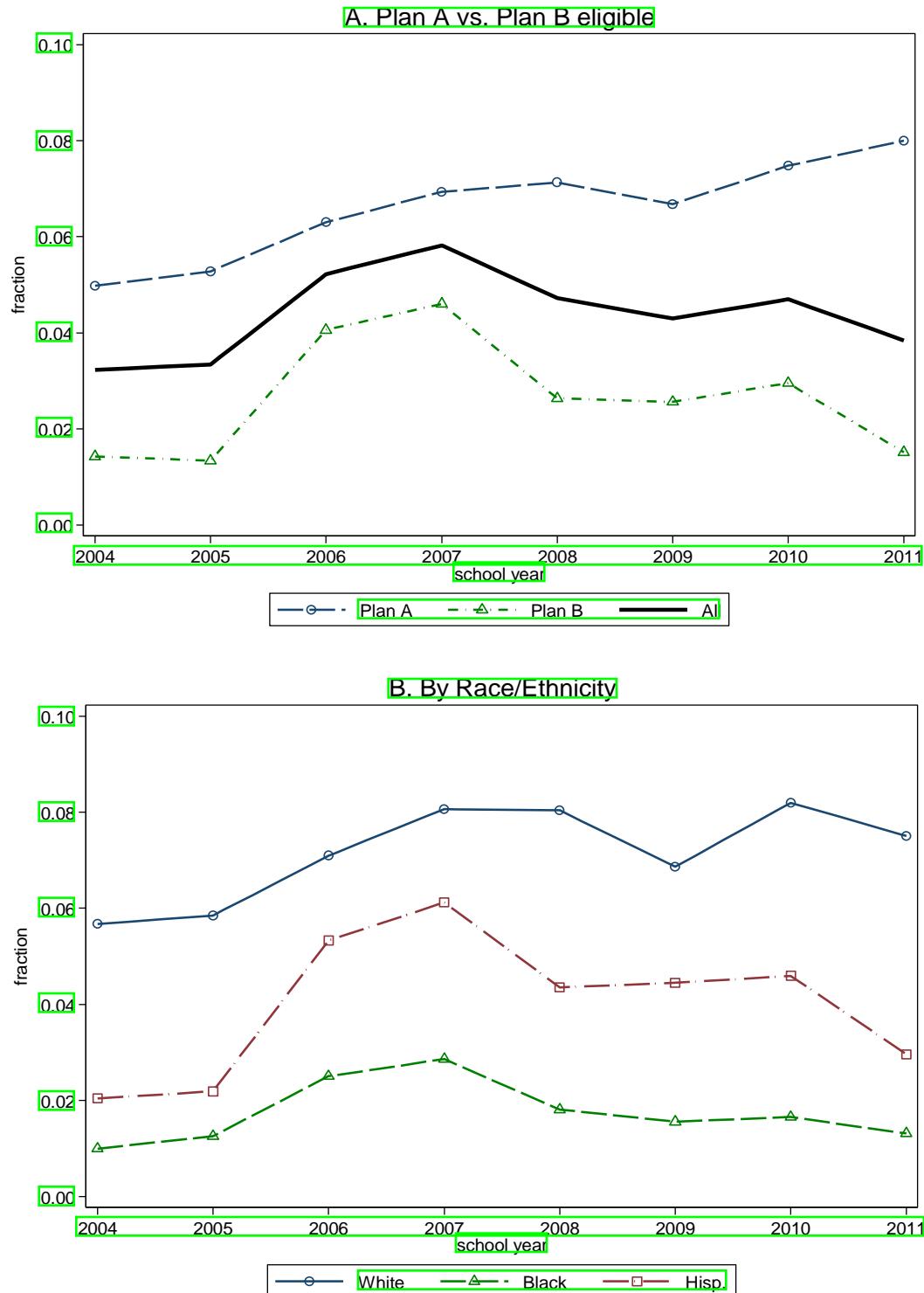


Table 3. Impact of Screening Program on Fraction of Gifted Third Graders

	Pre (2004-5)	Post (2006-7)	Difference	Trend-Adjusted Difference
	(1)	(2)	(3)	(4)
A. School/Grade/Year Aggregated Data -- Difference and Diff in Differences				
1. District Schools	0.036	0.053	0.017	--
	(0.004)	(0.005)	(0.004)	
2. District Schools Relative to Comparison Schools (p-score weighted)	0.003	0.019	0.016	0.016
	(0.005)	(0.006)	(0.003)	(0.003)
B. Student-Level Data -- Differences Design Only				
3. All Students (n=78,065 in 4 years)	0.033	0.055	0.022	0.015
			(0.001)	(0.003)
4. Plan A- Eligible Students (n=40,118 in 4 years)	0.051	0.066	0.015	0.006
			(0.002)	(0.005)
5. Plan B- Eligible Students (n=37,947 in 4 years)	0.014	0.043	0.029	0.025
			(0.002)	(0.004)
6. White Students (n=25,910 in 4 years)	0.058	0.076	0.018	0.007
			(0.003)	(0.007)
7. Black Students (n=27,108 in 4 years)	0.011	0.027	0.016	0.009
			(0.002)	(0.004)
7. Hispanic Students (n=19,855 in 4 years)	0.021	0.057	0.036	0.027
			(0.003)	(0.006)

Notes: Standard errors, clustered by school, in parentheses. Models in Panel A are based on publicly available school/grade/year data for schools in the District (row 1) or the District and a other districts in the same state (row 2). Models in Panel B are based on students in analysis sample for District schools only.

Appendix Figure 1. District Trends in Fraction Gifted by End of Third Grade



Appendix Table 1: Regression Discontinuity Estimates for Impact of Having NNAT Score Above Screening Threshold

	Dependent Variable:		
	<u>IQ Tested</u>	<u>IQ Tested and IQ Above Gifted</u>	<u>IQ Tested and Placed in Gifted Program</u>
		<u>Threshold</u>	
	(1)	(2)	(3)
<u>1. Plan A Eligible</u>	0.211 (0.026)	0.065 (0.018)	0.066 (0.016)
<u>number of obs.</u>	<u>9,444</u>	<u>9,444</u>	<u>9,444</u>
<u>2. Plan B Eligible</u>	0.522 (0.018)	0.120 (0.012)	0.068 (0.011)
<u>number of obs.</u>	<u>11,846</u>	<u>11,846</u>	<u>11,846</u>

Notes: standard errors in parentheses. Table reports estimated coefficient for dummy indicating NNAT score above screening threshold (130 points for Plan A group, 115 points for Plan B group) in model that also includes a constant. Dependent variables indicated in column heading. Sample includes students with NNAT scores within 20 points of the group-specific threshold.