

Proc Natl Acad Sci U S A. 2016 Nov 29; 113(48): 13678–13683.

Published online 2016 Nov 15. doi: [10.1073/pnas.1605043113](https://doi.org/10.1073/pnas.1605043113)

Economic Sciences

PMCID: PMC5137751

PMID: [27856741](https://pubmed.ncbi.nlm.nih.gov/27856741/)

## Universal screening increases the representation of low-income and minority students in gifted education

[David Card](#)<sup>a,1</sup> and [Laura Giuliano](#)<sup>b,1</sup>

<sup>a</sup>Center for Labor Economics, Department of Economics, University of California, Berkeley, CA, 94720-3880;

<sup>b</sup>Department of Economics, University of Miami, Coral Gables, FL, 33124

<sup>1</sup>To whom correspondence may be addressed. Email: [card@econ.berkeley.edu](mailto:card@econ.berkeley.edu) or [l.giuliano@miami.edu](mailto:l.giuliano@miami.edu).

Edited by Greg J. Duncan, University of California, Irvine, CA, and approved October 12, 2016 (received for review March 27, 2016)

Author contributions: D.C. and L.G. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

[Copyright notice](#)

### SIGNIFICANCE

[Go to:](#)  [Go to:](#)

A longstanding concern about gifted education in the United States is the underrepresentation of minorities and economically disadvantaged groups. One explanation for this gap is that standard processes for identifying gifted students, which are based largely on the referrals of parents and teachers, tend to miss many qualified students. Consistent with this hypothesis, we find that a universal screening program in a large urban school district led to significant increases in the numbers of poor and minority students who met the IQ standards for gifted status. Our findings raise the question of whether a systemic failure to identify qualified students from all backgrounds may help explain the broader pattern of minority underrepresentation in all advanced K–12 academic programs.

**Keywords:** gifted identification, universal screening, underrepresentation

### ABSTRACT

[Go to:](#)  [Go to:](#)

Low-income and minority students are substantially underrepresented in gifted education programs. The disparities persist despite efforts by many states and school districts to broaden participation through changes in their eligibility criteria. One explanation for the persistent gap is that standard processes for identifying gifted students, which are based largely on the referrals of parents and teachers, tend to miss qualified students from underrepresented groups. We study this hypothesis using the experiences of a large urban school district following the introduction of a universal screening program for second graders. Without any changes in the standards for gifted eligibility, the screening program led to large increases in the fractions of economically disadvantaged and minority students 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 were all systematically “underreferred” in the traditional parent/teacher referral system. Our findings suggest that parents and teachers often fail to recognize the potential of poor and minority students and those with limited English proficiency.

Low-income and minority students are substantially underrepresented in gifted and talented education programs in the United States ([1](#), [2](#)). In 2012, 7.6% of White K–12 students participated in gifted and talented programs nationwide, compared with only 3.6% of Blacks, 4.6% of Hispanics, and 1.8% of English learners ([ocrdata.ed.gov/StateNationalEstimations/Estimations\\_2011\\_12](http://ocrdata.ed.gov/StateNationalEstimations/Estimations_2011_12)). Some of this gap may be due to differences in measured cognitive ability of students from different backgrounds and biases in these measures. However, the standard processes for gifted screening are based on teacher and parent referrals, and there is evidence of underreferral of qualified students from disadvantaged backgrounds—suggesting that teacher/parent discretion in the referral process may be a further barrier ([3–7](#)). If so, then a comprehensive and objective screening program might be able to raise gifted participation rates among underserved groups by increasing their referral rates for gifted evaluation.

We test this hypothesis using data from a unique natural experiment conducted by a large and diverse school district in the state of Florida (hereafter “the District”). State law dictates that students must achieve a minimum of 130 points on a standard IQ test to qualify for gifted status. English language learners (ELLs) 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 mainly comprised White children from higher-income neighborhoods. Only 28% of gifted students in third grade were Black or Hispanic, compared with 60% of all students in the District. Thirteen regular elementary schools in the District had no gifted children in third grade in 2004 or 2005, but the gifted rate was nearly 10% at the 13 schools with the lowest fraction of FRL students.

In response to these disparities, the District introduced a universal screening program in spring 2005. Before this, candidates for gifted status were identified through parent and teacher referrals, mainly occurring in first and second grades. Under the new program, all second graders completed the Naglieri Non-Verbal Ability Test (NNAT), a nonverbal test intended to assess cognitive ability independent of linguistic and cultural background ([8](#)). The NNAT takes less than an hour to complete and was administered by teachers in the classroom. The NNAT scores were used to construct a nationally normed index with a mean of 100 and SD of 15, similar to a standard IQ test. All students scoring at least 130 points on the test, and ELL/FRL students scoring at least 115 points, were automatically eligible to be referred for full evaluation and regular IQ testing by District psychologists. Because students could still be nominated for testing by parents or teachers as in earlier years, the aim of the screening program was to supplement the traditional referral system and boost referral rates for underrepresented groups.

The other key features of the District’s gifted identification process remained unchanged. Referred students were placed in a queue for a full IQ test given by a District psychologist, although parents could bypass the queue by paying to have their child tested privately. Students with IQs above the relevant threshold were eligible for gifted status, with the final determination based on parent and teacher inputs and scores on a checklist of “gifted indicators.” ([Supporting Information](#) provides more details on the District’s gifted screening and identification procedures. See ref. [9](#) for additional information on the District's gifted program.) Importantly, the IQ thresholds and other requirements for gifted eligibility were unchanged. Any increase in the number of students identified as gifted following the introduction of the program can thus be attributed to the screening effort, and not to a relaxation of the standards for gifted status. [While the screening program may have raised parent and teacher awareness about the gifted program, the return of gifted rates to their prescreening levels after the program was suspended in 2011 ([Fig. 1](#)) suggests that increased awareness cannot explain the rise in

gifted rates after the program's introduction.]

[Open in a separate window](#)

[Fig. 1.](#)

Fraction gifted by end of third grade, District schools vs. matched comparison schools. (See [SI Materials and Methods](#) for details.)

As shown in [Fig. 1](#), comparisons across cohorts of third graders suggest the introduction of universal screening led to large increases 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 Florida districts was quite stable. However, because of financial pressures caused by the Great Recession, the District cut funding for IQ testing in 2007 and suspended the screening program in 2010. By 2011, the gifted share of third graders had returned to the level of 2004–2005. Meanwhile, program changes in other districts led to a gradual increase in gifted rates at the comparison schools after 2007.

In light of the history of the District's screening program and the striking patterns in [Fig. 1](#), we focus on simple “pre/post” comparisons between third graders in 2004–2005 (the two cohorts before the introduction of universal screening) and those in 2006–2007 (the two cohorts after). We confirm that the 2004–2005 (“pre”) cohorts form a valid comparison group for the 2006–2007 (“post”) cohorts. We then use between-cohort differences to measure the impact of the program on gifted participation rates, and to characterize the two key groups: students who were identified as gifted in the post cohort and would also 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. We refer to the former group as the “always takers” and the latter as the “compliers” [as in a standard analysis of experimental designs with incomplete compliance ([10](#))]. By studying the characteristics of the compliers and their distribution across schools, we gain insight into the types of students who would normally “fall through the cracks” of the traditional referral system.

Our analysis yields three main conclusions. First, the introduction of the screening program led to a large increase in the fraction of students classified as gifted. Second, the newly identified gifted students were disproportionately poor, Black, and Hispanic, and less likely to have parents whose primary language was English. They were also concentrated at schools with high shares of poor and minority students and low numbers of gifted students before the program. Thus, the experiences of the District confirm that a universal screening program can significantly broaden the diversity of students in gifted programs. Third, the distribution of IQ scores for the newly identified students was similar to the distribution for those identified under the old system, particularly among students who qualified under the Plan B eligibility standard. The newly identified group included many students with IQs well above the minimum eligibility threshold, implying that even high-ability students from disadvantaged groups were being overlooked under the traditional referral system.

## MATERIALS AND METHODS

[Go to:](#)  [Go to:](#)

**Student-Level Data and Sample Description.** We use deidentified longitudinal records of students

who were enrolled in the District for third grade between 2004 and 2011. (The research was approved by the Institutional Review Boards of the District and the National Bureau of Economic Research. Since we used only preexisting, deidentified records and did not conduct an experiment, we had no informed consent procedures.) For most of our analysis, we limit attention to students who were in third grade during the spring semester of the years from 2004 to 2007 and attended one of the 140 larger elementary schools in the District, excluding charter schools and other special schools.

[Table 1](#) presents descriptive information on the students in our sample, measured at the end of third grade. Data from the prescreening period (2004 and 2005) show the District's student body was racially diverse, with 35% White non-Hispanic students, 34% Black non-Hispanics, 25% Hispanics, and 3% Asians. Some 45% were eligible for FRL, and 11% were ELLs. Altogether, 49% of third graders were either FRL or ELL (or both) and were therefore eligible to be evaluated for giftedness under the state's Plan B standard (with a lower IQ threshold). The other 51% had to meet the regular (Plan A) eligibility requirements. [Because FRL and ELL status may change over time (e.g., as English learners transition out of the language program), a student can be referred for IQ testing and placed in the gifted program as Plan B eligible in second grade but recorded as Plan A eligible in third grade. We note in *Results* when changes in student status are important for interpreting our results.] Overall, about 16% of students in the prescreening cohorts had an IQ score on record by the end of third grade, and 3.3% were classified as gifted.

**Table 1.**

Characteristics of third-grade students at larger elementary schools, 2004–2005 versus 2006–2007

[Open in a separate window](#)

Sample is all first-time enrollees in third grade between 2004 and 2007 at 140 larger elementary schools in the District; 2004 refers to 2003–2004 school year.

\* Achievement is the average of reading and math scores on statewide tests, standardized across third graders in the District to have mean 0 and SD 1.

Comparing mean characteristics of nondisadvantaged, or Plan A, students with the Plan B eligible group, we see that White and Asian students are overrepresented in the Plan A group and Blacks and Hispanics are overrepresented in the Plan B group. Consistent with the ELL status of many Plan B eligibles, these students are less likely to have parents whose primary language is English. Plan B students also had lower average achievement (measured as the average of reading and math scores on statewide tests in third grade), and they attended schools with higher fractions of minority and FRL participants, reflecting the District's residential segregation patterns and neighborhood-based school assignments. Importantly, despite facing a lower threshold for gifted status, Plan B students were less likely than Plan A students to have an IQ score on record and were much less likely to be classified as gifted.

**Analysis of Student-Level Data.** We use a simple pre/post differences or interrupted time series

approach to analyze the student-level data from the District. For third graders as a whole and for subgroups, we present estimates from comparisons of the gifted share between the 2004–2005 (pre) and 2006–2007 (post) cohorts, and from models that adjust for a time trend. We use the same data to analyze the characteristics of the students who are impacted by the program (the compliers).

Controlling for a smooth trend (reflecting longer-run demographic factors), any change in the mean characteristics of the gifted population can be attributed to disparities in the impact of the screening program across students with different characteristics. We estimate the fraction of compliers with a given characteristic as the trend-adjusted change in the population share that is both gifted and has the characteristic, divided by the trend-adjusted change in the share that is gifted (see [SI Materials and Methods](#)). We then compare the characteristics of compliers to those of the always takers who were identified under the traditional referral system. SEs are adjusted to allow for school-level error components in the fraction of gifted students.

The validity of the interrupted time series design is supported by evidence, presented in [Table 1](#), that, among District third graders as a whole, students in the two post cohorts are very similar to those in the two pre cohorts. The only notable difference is a small decline in the fraction of Whites, offset by small rises in the fractions of Blacks and Hispanics. This stability suggests that student outcomes in the pre cohorts (2004–2005) represent a plausible counterfactual for the outcomes of students in the post cohorts (2006–2007), particularly if we adjust for a trend to capture long-run demographic shifts. Our approach is further supported by the patterns in [Fig. 1](#), which compares the evolution of gifted participation rates for third graders in the District and in a matched comparison group of schools from other districts in Florida. Between 2004 and 2007, the gifted share of third graders in the comparison group is relatively stable, with only a slight upward trend (3.5% in 2004, 3.6% in 2005, and 3.7% in 2006 and 2007).

**Use of School/Grade-Level Data from Other Districts.** [Fig. 1](#) is constructed using state data on 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 use larger, noncharter elementary schools ( $\geq 40$  third graders in each year). To construct a comparison group of schools from other districts that closely matches the characteristics of District schools, we weight the data using the estimated probability that a school with given characteristics is found in the District (see [SI Materials and Methods](#)).

We also use these school-level data to construct an alternative, “difference-in-difference” estimate of the impact of the screening program on gifted participation rates. Here the matched comparison group provides a counterfactual trend for gifted rates in the District, and we estimate the pre–post difference in the District’s gifted rate net of the rate in the comparison group. As in the student-level analysis, we estimate models that allow for trends and SEs that allow for school-level clustering.

Because our experiment is confined to a single district, the precision of our estimates could be overstated by the presence of unobserved district-wide factors that cause the gifted rate to vary from year to year. To assess this concern, we constructed the empirical distribution of changes in the average third-grade gifted rate from 2004–2005 to 2006–2007 for the 23 larger districts in Florida (those with  $\geq 20$  elementary schools), excluding one other district that adopted a screening program in 2005, but including the District (which, with 140 elementary schools, is one of the largest in the state). We show this distribution in [Fig. 2](#), along with the confidence intervals associated with the changes in each district (based on the variability in changes in the gifted rate across schools in the district). Two features are clear. First, among other larger districts, the variability in the average 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. The empirical mean and SD of the changes across the 22 other districts are  $-0.001$  and  $0.007$ . The observed change in the District is 2.65 SDs from the mean for the other districts, which can be interpreted as a  $t$  test statistic. 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.

[Open in a separate window](#)

[Fig. 2.](#)

Cumulative distribution of changes in the fraction of gifted third-grade students between 2004–2005 and 2006–2007. Districts are in the same state, with 20+ elementary schools. Lines represent 95% confidence interval.

## RESULTS

Go to:  Go to:

**Impact of Screening on Rates of IQ Testing.** [Table 1](#) shows differences between the pre and post cohorts in the fraction of students who are IQ tested by the end of third grade, the mean IQ score of those tested, and the fraction who are tested and have an IQ score above the gifted eligibility cutoff. Students are IQ-tested for a variety of reasons, including assessment for intellectual and learning disabilities, so the number with an IQ test in each cohort exceeds the number evaluated for gifted status, and their average IQ score is well below the gifted eligibility threshold. Between the pre and post cohorts, however, the increase in IQ testing was presumably driven by a rise in referrals caused by the screening program (see [SI Results](#), [Fig. S1](#)). With the introduction of universal screening, the overall fraction of third graders with an IQ test rose by eight percentage points (ppt.), with a much larger gain for Plan B eligibles—large enough to close the testing gap between the two groups. The fraction of third graders with IQs above the gifted threshold also rose in both groups, again with a larger increase among Plan B eligible students. In the post cohorts, 6.4% of the Plan B group had IQs above the threshold, compared with 7.7% of Plan A eligibles. Most importantly, the fraction of students classified as gifted also increased in both groups, rising from 5.1% to 6.6% among Plan A students and from 1.4% to 4.3% in the Plan B group.

[Open in a separate window](#)

[Fig. S1.](#)

Frequency distribution of IQ scores, pre and post universal screening.

**Impact of Screening on Gifted Participation Rates.** [Table 2](#) presents estimates of the impact of the screening program on gifted placement rates from different models and for various subgroups. The first row shows the difference-in-difference estimates based on the school-level data. The simple difference

of differences between District and comparison group schools is 1.6 ppt., with an SE of 0.3 ppt. The estimate from a model that allows gifted rates at schools in the state to have a trend over time is virtually the same.

### Table 2.

Estimates of screening program impact on gifted share of third graders

[Open in a separate window](#)

First row shows sample means in the pre and post years of the gap in the fraction gifted between District schools and comparison group schools in other districts, and estimates of the program impact based on the difference in differences. Remaining rows show sample means and linear probability estimates (models 1 and 2) or odds ratios (model 3) using students in District schools only (see [Table 1](#) for sample sizes). Parentheses contain SEs, clustered by school in all models. (\*\* $P < 0.01$ ; \* $P < 0.05$ .)

The remaining estimates in [Table 2](#) are from models that use student-level data from the District; [Fig. 3](#) shows the time series patterns in gifted rates based on these data. The simple difference estimate for all District third graders shows an increase of 2.2 ppt. The trend-adjusted difference (model 2) is 1.5 ppt. with an SE of 0.3 ppt., which is nearly identical to the difference-in-differences estimate. The final column (model 3) shows odds ratios estimated using a logit model that adjusts for a trend as in model 2. The overall impact for third graders translates into a 45% increase in the odds of being identified as gifted.

[Open in a separate window](#)

### [Fig. 3.](#)

District trends in fraction gifted by end of third grade.

Because the gifted shares differed dramatically across subgroups before universal screening, the odds ratios from model 3 are particularly helpful when comparing impacts for different groups. For Plan B students, the estimate implies a 174% increase in the odds of being identified as gifted—much larger than the Plan A increase of 11%. The impacts for Blacks and Hispanics are also very large: The odds rose by 74% for Blacks and by 118% for Hispanics (the difference between these two estimates is not significant;  $P = 0.44$ .) By contrast, the impact for Whites is relatively small (12%) and differs significantly from the impact for Blacks and Hispanics combined ( $P < 0.01$ ).

**Characteristics of Newly Identified Gifted Students.** The evidence in [Table 2](#) suggests that the changes in gifted participation induced by the screening program were very different across economic and racial groups. [Table 3](#) 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 as gifted under the traditional referral system (the always takers). The first column shows the mean characteristics of all gifted students in the pre cohorts; by definition, this group consists only of always takers. The second column shows the characteristics of all gifted students in the post cohorts; this group includes both always takers and compliers. The estimated characteristics of the compliers are shown in the third column. The fourth column shows estimates of the differences in mean characteristics between the compliers and the always takers.

### Table 3.

Characteristics of all gifted third-grade students and those identified through screening

[Open in a separate window](#)

For full sample,  $n = 78,065$ ; for Plan B sample,  $n = 37,947$ . Parentheses contain SEs, clustered by school. (\*\* $P < 0.01$ ; \* $P < 0.05$ .)

\* Achievement is the average of reading and math scores on statewide tests, standardized across third graders in the District to have mean 0 and SD 1.

This analysis shows that about 80% of the compliers are Plan B eligible. Compared with the always takers, the compliers are disproportionately Black (23% vs. 12%) and Hispanic (46% vs. 16%), and they are substantially less likely to be White (8% vs. 61%) or to have English-speaking parents (29% vs. 74%). They also attended schools with relatively high fractions of FRL participants and non-White students, and they are somewhat more likely to be female (56% vs. 45%). [In the final column of [Table 3](#), which shows the estimated differences in complier and always-taker characteristics, the estimates for the Black share (0.12) and the Hispanic share (0.30) differ significantly from the White share estimate of  $-0.53$  ( $P < 0.01$ ) but do not differ significantly from each other ( $P = 0.25$ ). Additional subgroup analyses are reported in [SI Results](#) and [Table S1](#).] Finally, the compliers have about 0.4 SD units lower third-grade test scores than the always takers, and IQ scores that are about 0.5 SD units lower. In interpreting the IQ gap, however, it is important to keep in mind that most compliers are Plan B eligible, and are therefore subject to a lower IQ threshold than the Plan A group who comprise most of the always takers.

### Table S1.

Detailed characteristics of all gifted third-grade students and those identified through screening

[Open in a separate window](#)

Number of cases  $n = 78,065$ . (\*\* $P < 0.01$ ; \* $P < 0.05$ ; † $P < 0.10$ .)

Given this fact, [Table 3](#) also reports a parallel analysis for the Plan B subpopulation. The characteristics of the Plan B compliers are quite similar to those of the overall complier group, reflecting the fact that 80% of all compliers are Plan B eligible. The Plan B compliers are broadly similar to the Plan B always takers, but are 21 ppt. more likely to be Hispanic, and 27 ppt. less likely to have parents who speak English. The latter gaps suggest that language may be an important barrier to the identification of qualified gifted children in a referral-based system.

Looking at the last two rows of [Table 3](#) for the Plan B group, the compliers have lower achievement scores than the always takers, but about the same mean IQ scores. This suggests that the traditional referral system tends to miss disadvantaged students with modest achievement levels, regardless of their cognitive abilities. Teachers and parents may simply not recognize the abilities of many of these students. Further insights into the cognitive abilities of the compliers are provided in [Fig. 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. In the Plan B group, the two distributions are similar—although the compliers are somewhat less likely to have scores very close to the 116 cutoff. Moreover, a full 20% of the compliers have IQs of 130 or higher (vs. 25% of the always takers). This suggests that many high-ability disadvantaged students are at risk for being overlooked in a traditional parent/teacher referral system.

[Open in a separate window](#)

[Fig. 4.](#)

IQ distributions of always takers and students identified by screening.

The distributions of scores in the Plan A group are also interesting. About 30% of the compliers have scores under 125 points, vs. 8% of the always takers. Virtually all those with  $IQ < 125$  were classified as Plan B at the time of IQ testing but are Plan A by the end of third grade (mainly because they transition out of ELL status). However, apart from this group, many of the newly identified Plan A gifted students had IQ scores well above the minimum 130-point threshold. We thus conclude that the traditional referral system also misses some high-ability nondisadvantaged students.

**Impact on School Distribution.** A key feature of the District's gifted program before universal screening was the unequal distribution of gifted students across schools. This is illustrated by the blue line in [Fig. 5](#), which plots the cumulative share of gifted third graders in the 2004–2005 cohorts among the District's 140 larger elementary schools against the cumulative share of all third-grade students. The schools are ranked by their fraction gifted, with the first school contributing the largest relative share of gifted students. Half of gifted students in the pre cohort were at schools that enrolled only 18% of the entire third-grade population, whereas half of all third-grade students in the District were at schools that enrolled a total of only 16% of the gifted students.

[Open in a separate window](#)

The impact of the screening program is confirmed by the green line in [Fig. 5](#), which shows the same plot for the 2006–2007 cohorts. This line is much closer to the 45° line, implying a more equal distribution of gifted students across schools. In particular, all 140 larger elementary schools in the District had at least one gifted third-grade student in 2006 or 2007, whereas, in 2004 and 2005, 13 schools had no gifted students. The third (red) line in [Fig. 5](#) shows the distribution of the compliers across schools. This line is mostly below the 45° line, implying that, on average, the compliers were likely to come from schools with relatively low fractions of gifted students in the pre cohort.

## DISCUSSION

Go to:  Go to:

Critics of gifted education programs have long noted the underrepresentation of minorities, nonnative 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 ([11–13](#)). Over the past 30 y, states and school districts have introduced new criteria for determining giftedness (e.g., grids using a combination of IQ and achievement) and also adopted different thresholds for minorities and economically disadvantaged groups ([14](#)). 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 representation gap is that the referral processes by which students are nominated for gifted evaluation tend to systematically miss many qualified minorities and economically 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 174% increase in the odds of being identified as gifted among all disadvantaged students, with a 118% increase for Hispanics and a 74% increase for Blacks.

A comparison of the newly identified gifted students to those who would have been identified even without screening shows that Black and Hispanic students, FRL participants, ELLs, and girls were all systematically “underreferred” to the gifted program. Newly identified gifted students were more likely to come from schools in poor neighborhoods with 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, although they had lower standardized achievement scores. We hypothesize that parents and teachers often fail to recognize the potential of many poor, immigrant, and minority children with less than stellar achievement levels. The large impacts for Hispanics, ELLs, and students with non-English speaking parents suggest that language differences, in particular, are an important barrier to identifying gifted students. At the same time, the combination of large impacts for Blacks and negligible effects for Whites suggests that factors related to race/culture may also play a role. Thus, although our findings show that underrepresented groups fare better under a screening process that places less weight on subjective assessments, they also suggest that training to improve the intercultural competence of teachers may be beneficial. (It is also worth noting that while biases or disparate impact of IQ tests are a common concern, our findings are consistent with the argument often made by proponents of IQ

testing that it can serve to limit biases associated with subjective judgment. We thank a referee for this point.)

An important limitation of our analysis is that it pertains to only a single school district. Although the student population in the District is highly diverse, and arguably representative of the student population in many large urban districts, the District's gifted policies have a number of distinctive features. First, both the use of a nonverbal screening test and the lower Plan B referral threshold may have been important factors in the success of the District's screening program. Second, the District has strict IQ thresholds for eligibility, and, before the introduction of the screening program, the District had relatively low gifted rates (around 3.5% for third graders) that reflected the importance of the IQ thresholds. At a minimum, however, our findings suggest that the underrepresentation 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. More broadly, disadvantaged groups are significantly underrepresented in advanced academic programs at all levels of K–12 education (15). Our findings raise the question of whether this larger pattern of underrepresentation is also due in part to a failure to identify and serve capable students from all backgrounds.

## SI MATERIALS AND METHODS

Go to:  Go to:

**Method for Estimating Complier Characteristics.** Define  $g_\tau$  as the fraction of all third-grade students who are gifted in period  $\tau$ , and define  $\bar{x}_\tau$  as the mean value of characteristic  $x$  for these gifted students. Let  $\tau = 1$  represent the two preprogram years just before the introduction of universal screening (2004–2005) and  $\tau = 2$  represent the two post years (2006–2007). Further, define  $g_c$  as the fraction of all students in the postscreening period who are gifted compliers (students identified as gifted through the screening process who would not have been identified in the prescreening period), and define  $\bar{x}_c$  as the mean characteristic of these compliers.

Note that, by definition, all gifted students in period 1 are always takers, whereas the population of gifted students in period 2 includes both always takers and compliers. Abstracting from demographic trends and changes over time in the population share of always takers, we have  $g_2 = g_1 + g_c$ . Likewise, the mean characteristic of gifted students in period 2 can be decomposed as follows:

$$\bar{x}_2 = \frac{g_1}{g_1 + g_c} \bar{x}_1 + \frac{g_c}{g_1 + g_c} \bar{x}_c .$$

From this we can infer

$$\bar{x}_c = \frac{g_1 + g_c}{g_c} \bar{x}_2 - \frac{g_1}{g_c} \bar{x}_1$$

or

$$\bar{x}_c = \frac{g_2\bar{x}_2 - g_1\bar{x}_1}{g_2 - g_1}.$$

We use this expression to characterize the compliers and compare them to the always takers with mean characteristics  $\bar{x}_1$ .

To construct an estimate of  $\bar{x}_c$  for each student characteristic  $x$ , we fit the following models using student-level data from our sample of District third graders in 2004–2007:

$$g_{it} = \alpha_0 + \alpha_1 \cdot year_t + \gamma \cdot Post_t + \varepsilon_{it} \quad [S1]$$

$$w_{it} = \beta_0 + \beta_1 \cdot year_t + \delta \cdot Post_t + \mu_{it}. \quad [S2]$$

Here  $g_{it}$  is a gifted indicator for student  $i$  in the third-grade cohort for year  $t$ ;  $w_{it} = g_{it} \cdot x_{it}$ ;  $year_t$  is a trend variable, and  $Post_t$  is an indicator for the postscreening period. From [Eq. S1](#), the estimated coefficient  $\hat{\gamma}$  provides a (trend-adjusted) estimate of  $g_2 - g_1$ , the impact of the screening program on the fraction of students that are gifted. From [Eq. S2](#), the coefficient  $\hat{\delta}$  provides a trend-adjusted estimate of  $g_2\bar{x}_2 - g_1\bar{x}_1$ . Hence we can estimate  $\bar{x}_c$  as  $\hat{\delta}/\hat{\gamma}$ . To calculate the SE, we estimate [Eqs. S1](#) and [S2](#) simultaneously using Stata's `suest` command, and we allow for clustering by school.

**Method for Constructing Comparison Group of Schools.** To ensure that the comparison of gifted rates in [Fig. 1](#) and the results from our difference-in-difference models are not driven by differences in demographic trends between District and non-District schools, we use inverse probability (propensity score) weighting to construct the comparison group of schools. First, we fit a propensity score model to the pooled sample of District schools and larger elementary schools in other districts in the state, excluding one other district that adopted a screening program in 2005. This sample includes a total of 1,397 noncharter elementary schools with at least 40 third graders in each year. We estimate a logit model for the probability that school  $s$  is a District school,

$$\Pr(District_s = 1 | X_s) = F(\beta X_s),$$

where  $F$  is the cumulative logistic function and  $X$  includes the following school-level variables: the mean fraction gifted in 2004–2005; the mean FRL participation rate of third graders in 2004–2005, in 2006–2008, and in 2009–2011; the mean fraction of Blacks and Hispanics among third graders in 2004–2005, 2006–2008, and 2009–2011; and the mean number of third-grade students in all years from 2004 to 2011. We then use the coefficient estimates to construct a predicted probability (propensity score)  $\hat{p}_s$  for each school, and we calculate the weight for non-District school  $s$  as  $1 / (1 - \hat{p}_s)$ .

In addition to the propensity scores weighting, the school-level data are also weighted by mean school enrollment in all years. We use similar weights when estimating models using school-level data.

## SI DISTRICT'S GIFTED SCREENING AND IDENTIFICATION PROCEDURES

Go to:  Go to:

The District's process for identifying gifted students involves two steps. First, students must be identified as potentially gifted and referred for further evaluation. Referred students then receive an evaluation that includes a standard IQ test, which must be administered individually by a licensed

psychologist or school psychologist. (Individual IQ tests used by the District include the Wechsler Intelligence Scale for Children, the Stanford–Binet IQ test, and the Leiter International Performance Scale.) Students with IQs above a threshold are eligible for gifted status, 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 learning, motivation, leadership, creativity, and adaptability. (The eligibility checklist for Plan B students also includes measures of academic achievement and family background factors.)

The IQ thresholds for gifted eligibility are largely determined by state law. Specifically, state law dictates that students who are neither ELLs nor participants in the FRL program must achieve a minimum of 130 points on a standard IQ test to qualify for gifted status. To address potential biases in these tests and their disparate impact on non-English speakers and economically disadvantaged groups, state law allows districts to set a lower threshold, known as Plan B eligibility, for ELLs and FRL participants. [For nondisadvantaged students (who face a 130 IQ threshold), state law allows those scoring within a standard error (three points) of the threshold to be placed in the gifted program if there are extenuating circumstances. A small fraction of gifted students have scores one, two, or three points below the 130 threshold. For disadvantaged (Plan B) students, however, no exceptions are made.] Like many other districts in the state, the studied District set a threshold of 116 for Plan B eligible students.

The District offers IQ testing by a school psychologist free of charge to students who are referred for gifted evaluation, and students are placed in a queue for testing upon referral. [Students may also be referred for free District IQ testing (or seek private testing) for other reasons, including evaluation for learning disabilities, intellectual disabilities, and other special needs. See [SI Results](#) for details.]

Alternatively, parents can bypass the queue by paying to have their child tested privately and submitting the IQ score to the District. There is a thriving private market for IQ testing in the District, with many psychologists offering first-time IQ tests as well as retesting for children who fail to meet the state mandated standards on an earlier test. Advertisements posted by these psychologists suggest that a typical price for an evaluation and IQ test was \$500 to \$1,000 in 2013 and that their target market is mainly nondisadvantaged families whose children have to meet the 130-point IQ threshold (known as the Plan A eligibility rule). Indeed, an analysis of the testing records for third graders in 2005 and 2006 shows that 20% of Plan A eligible students qualified for gifted status on the basis of a private IQ test. In contrast, less than 1% of Plan B eligible students qualified through a private test.

Before the introduction of the screening program, candidates for gifted status in the District were referred for IQ testing through a process that relied on parent and teacher nominations. Referrals were concentrated among first and second graders to ensure that qualified students could begin receiving gifted services by third grade. (More information on the District's gifted program is presented in ref. 9. 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.) Under the universal screening program, introduced in spring 2005, all second graders in the District completed the NNAT. The NNAT—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. The scores from this test were rescaled to have a mean of 100 and an SD of 15, similar to a standard IQ test. (Specifically, the raw scores were combined with student age to construct a “nonverbal ability index” that is nationally normed to have a mean of 100 and a standard deviation of 15.) Nondisadvantaged students scoring at least 130 points on the test, and ELL/FRL students scoring at least 115 points, were eligible for referral for IQ testing. [While the 130 Plan A cutoff applied to both the NNAT test and the IQ test, the Plan B

NNAT cutoff (115) was one point lower than the Plan B IQ cutoff of 116. Also, 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.] 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, an analysis of records for third graders in 2006 and 2007 shows that most gifted students completed an IQ test within 6 mo of the screening test.

The other key features of the District's gifted identification process remained unchanged after the introduction of universal screening. As in earlier years, students could still be nominated for IQ testing by parents or teachers, and parents could still bypass the queue for a free evaluation with a school psychologist by having their children tested privately. Importantly, the IQ testing procedures and eligibility cutoffs also remained unchanged. The screening program was therefore intended to supplement the traditional referral system and help narrow the gaps in referral rates between students from different backgrounds.

A key issue for the screening program was the need to expand 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, with a similar number in 2006. Because each test takes ~1 h to 1.5 h, the cost in overtime payments for testing staff was relatively large. In response to a budget crisis in 2007, the District elected to cut the overtime budget for testing staff. The result, clearly evident in [Fig. 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 in 2010.

## SI RESULTS

[Go to:](#)  [Go to:](#)

**The Impact of Universal Screening on Gifted and Nongifted Evaluations.** The results presented in [Table 1](#) show that rates of IQ testing among third graders, mean IQ scores among those tested, and the fraction of students with an IQ score above the eligibility threshold all increased between the pre and post cohorts. These changes are consistent with an increase in the number of IQ tests performed for the purpose of gifted evaluation. However, students may be IQ-tested for a variety of reasons, including assessment for learning disabilities or intellectual disabilities (defined as having an IQ below 70), and the fact that average IQ scores are well below the gifted eligibility thresholds in both the pre and post cohorts suggests that many students were referred for reasons other than gifted evaluation, even under universal screening.

Because our data do not include information on the reason an IQ test is performed, we cannot directly estimate the impact of universal screening on the number of referrals for gifted evaluation. However, it seems likely that the expansion in IQ testing following the introduction of the program in 2005 was driven entirely by an increase in gifted referrals, and that nongifted referrals and evaluations were unaffected by universal screening. First, the rate at which students were identified as either learning-disabled or intellectually disabled remained stable over the years of our analysis; roughly 3% of Plan A students and 5% of Plan B students were labeled as having one of these exceptionalities in both the prescreening and postscreening cohorts.

Second, a closer examination of the IQ distributions before and after universal screening provides further evidence that the program had no impact on nongifted referrals. [Fig. S1](#) compares the frequency

distribution of IQ scores for Plan A and Plan B eligible students in the prescreening and postscreening cohorts.\* Among both the Plan A and Plan B groups, there was little change in the numbers of students who took an IQ test and scored in the range of around 90 or below. In particular, roughly 1,230 Plan B students with an IQ test had IQ scores of 90 or below in both the pre and post cohorts. By contrast, there were substantial increases in the numbers of students scoring between 90 and 140, with especially large increases among Plan B eligible students.

**Detailed Characteristics of Newly Identified Gifted Students.** [Table S1](#) presents additional results from our analysis of gifted student characteristics. Similar to [Table 3](#), it presents a comparison of the newly identified gifted students—the compliers whose gifted status was changed by the screening program—to the students who would have been identified as gifted under the traditional referral system (the always takers). As in [Table 3](#), the final column shows estimates of the differences in mean characteristics between the compliers and the always takers. Here we provide a more detailed analysis of student demographics by examining the interactions of gender with race and ethnicity, of gender with FRL status, and of race and ethnicity with FRL status. We also examine achievement scores disaggregated by subject (math and reading). The results in [Table S1](#) suggest that, among subgroups defined by race/ethnicity and FRL status, White students who are not FRL participants are the least affected by universal screening, whereas Hispanic FRL participants are the most affected. The results also suggest that the gender difference in the impact of screening is present among White and Hispanic students (with girls affected more than boys) but not among Black students. (However, the gender difference is not statistically significant for any race group.) Finally, we find that students who were identified through the screening program have lower achievement test scores in both reading and math. The differential is somewhat larger in reading, but not significantly so.

## ACKNOWLEDGMENTS

[Go to:](#)  [Go to:](#)

We thank Cynthia Park, Jacalyn Schulman, and Donna Turner for their help in accessing and interpreting the data, and Attila Lindner, Carl Nadler, and Sydnee Caldwell for their expert assistance. Special thanks go to Hedvig Horvath for her input at many stages. The research reported here was supported by the Institute of Education Sciences, US Department of Education, through Grant R305D110019 to the National Bureau of Economic Research.

## FOOTNOTES

[Go to:](#)  [Go to:](#)

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

\*The notable dip in Plan A IQ scores in the 125 to 129 range and subsequent spike in the 130 to 135 range are likely due to a combination of manipulation and selective reporting of scores from private psychologists. State law allowing students to be considered for gifted status if they score between 127 and 129 when there are extenuating circumstances (see *SI District's Gifted Screening and Identification Procedures*) may put pressure on school psychologists to “top up” scores in this range to avoid appeals from parents. The market for private testing also likely contributes to this phenomenon. However, some of the “missing” scores just below 130 may be due to a failure to report low scores from a private evaluation to the District rather than to score manipulation. For an analysis of the extent of IQ score manipulation, see ref. [9](#).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1605043113/-/DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1605043113/-/DCSupplemental).

## REFERENCES

[Go to:](#)  [Go to:](#)

1. Ford DY. The underrepresentation of minority students in gifted education: Problems and promises

- in recruitment and retention. *J Spec Educ.* 1998;32(1):4–14.
2. Donovan MS, Cross CT. *Minority Students in Special and Gifted Education.* Natl Acad; Washington, DC: 2002.
  3. Woods SB, Achey VH. Successful identification of gifted racial/ethnic group students without changing classification requirements. *Roeper Rev.* 1990;13(1):1–26.
  4. Figlio D. 2005. *Names, Expectations, and the Black-White Test Score Gap* (Natl Bur Econ Res, Cambridge, MA), Work Pap 11195.
  5. Elhoweris H, Mutua K, Alsheikh N, Holloway P. Effects of children’s ethnicity on teachers’ referral and recommendation decisions in gifted and talented education. *Remedial Spec Educ.* 2005;26(1):25–31.
  6. Grissom JA, Redding C. Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open.* 2016;2(1):10.1177/2332858415622175.
  7. Ford DY, Grantham TC, Whiting GW. Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Except Child.* 2008;74(3):289–306.
  8. Naglieri JA, Ford DY. Assessing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT) *Gift Child Q.* 2003;47(2):155–160.
  9. Card D, Giuliano L. 2014. *Does Gifted Education Work? For Which Students?* (Natl Bur Econ Res, Cambridge, MA), Work Pap 20453.
  10. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica.* 1994;62(2):467–475.
  11. Borland JH, Wright L. Identifying young, potentially gifted, economically disadvantaged students. *Gift Child Q.* 1994;38(4):164–171.
  12. Renzulli JS. What makes giftedness? Re-examining a definition. *Phi Delta Kappan.* 1978;60(3):180–184.
  13. VanTassel-Baska J, Feng AX, Evans B. Patterns of identification and performance among gifted students identified through performance tasks: A three year analysis. *Gift Child Q.* 2007;51(3):218–231.
  14. McClain M-C, Pfeiffer S. Identification of gifted students in the United States today: A Look at state definitions, policies and practices. *J Appl Sch Psychol.* 2012;28(1):59–88.
  15. US Department of Education, Office for Civil Rights 2014. *Data Snapshot: College and Career Readiness* (US Dep Educ, Washington, DC), Issue Brief 3.