Nurturing for a Bright Tomorrow: Experimental Evidence for Two Kindergarten Cohorts

Authors

Matthew Lenard (WCPSS Contractor and corresponding author)

Angel Harris (Duke University)

Darryl Hill (former WCPSS employee)

Acknowledgements

This research is supported by a grant from the Jacob K. Javits Gifted and Talented Students Education Program (2014; PR/Award # S206A140032) to Duke University. The authors thank the project implementation team, which consists of staff from both Wake County Public School System and Duke University. They include current and former members of the school district's Academically or Intellectually Gifted (AIG) department, notably Shani Brown, Michelle Gainey, Melba Spivey, Alena Treat and staff from Duke's Social Science Research Institute (SSRI), notably Jackie Terrell and Kamilah Legette. Thanks to Margaret Gayle and Dan Turner for their contributions to earlier phases of this project and staff from Columbia University's Center for Benefit-Cost Studies of Education for valuable feedback during the final year of implementation.

October 2019

DRA Report No. 19.07

WAKE COUNTY PUBLIC SCHOOL SYSTEM

Abstract

In response to chronically low gifted identification rates in the majority of elementary schools among Black and Hispanic/Latino students, the Wake County Public School System (WCPSS) launched Nurturing for a Bright Tomorrow (NBT) in Fall 2014. NBT was a multi-component curricular enhancement program designed to train early elementary school teachers (K-2) to develop the skills and expectations required to help children attain gifted identification in elementary school. Results show that students enrolled in NBT schools, especially Hispanic/Latino students, outperformed their non-NBT counterparts on the Naglieri Nonverbal Ability Test (NNAT) and had correspondingly greater odds of meeting gifted thresholds on the NNAT. However, the odds that NBT students would qualify for the district's gifted program by grade 3 were generally not different than the odds for control group students. Thus, WCPSS should continue to pilot programs and policies with the potential to increase overall gifted identification rates and narrow racial identification gaps in schools with historically low gifted rates.

Introduction

In response to chronically low gifted identification rates in the majority of elementary schools among Black and Hispanic/Latino students, the Wake County Public School System (WCPSS) launched Nurturing for a Bright Tomorrow (NBT) in Fall 2014. NBT was a multi-component curricular enhancement program designed to train early elementary school teachers (K-2) to develop the skills and expectations required to help children attain gifted identification in elementary school. To achieve this aim, a joint implementation team consisting of WCPSS's Academically or Intellectually Gifted (AIG) department and Duke University partnered to train teachers, distribute curricular materials, and monitor implementation. NBT, through random assignment, was launched in 16 elementary schools for two cohorts of participants in Fall 2014 and Fall 2015.

This evaluation measures the impact of NBT on gifted identification indicators for program participants who were enrolled in an NBT school as kindergarteners and who were ultimately screened for gifted identification when they reached the third grade. Two cohorts passed through the intervention: (1) Fall 2014 kindergarten enrollees assessed for gifted identification in Fall 2017 and (2) Fall 2015 kindergarten enrollees assessed for gifted identification in Fall 2018. The causal impacts of NBT show that:

- Students enrolled in NBT schools in grades 1 and 2 outperformed their non-NBT counterparts on the NNAT test by roughly 0.07-0.14 standard deviations. The effect was even larger—sometimes double—for Hispanic/Latino students.
- Students enrolled in NBT schools in grades 1 and 2 had slightly higher odds of being identified as gifted based on informal NNAT thresholds compared with their control group counterparts.
- The odds that students enrolled in NBT would be identified as gifted based on district criteria generally did not differ from those odds for students in the control group.

Since NBT was randomly assigned, this evaluation represents a "gold standard" research design and the results can be interpreted as causal. Table 1 summarizes the main categories of research designs used by WCPSS and the conclusions that can be drawn from them.

Table 1
Nature of the Data Provided and Valid Uses

Research Design	Conclusions that Can be Drawn
☑ Experimental	We can conclude that the program or policy caused changes in outcomes because the research design used random assignment.
☐ Quasi-Experimental	We can reasonably conclude that the program or policy caused changes in outcomes because an appropriate comparison strategy was used.
☑ Descriptive ☑ Quantitative ☐ Qualitative	These designs provide outcome data for the program or policy, but differences cannot be attributed directly to it due to lack of a comparative control group.

Sources: List, Sadoff, & Wagner (2011) and What Works Clearinghouse (2014).

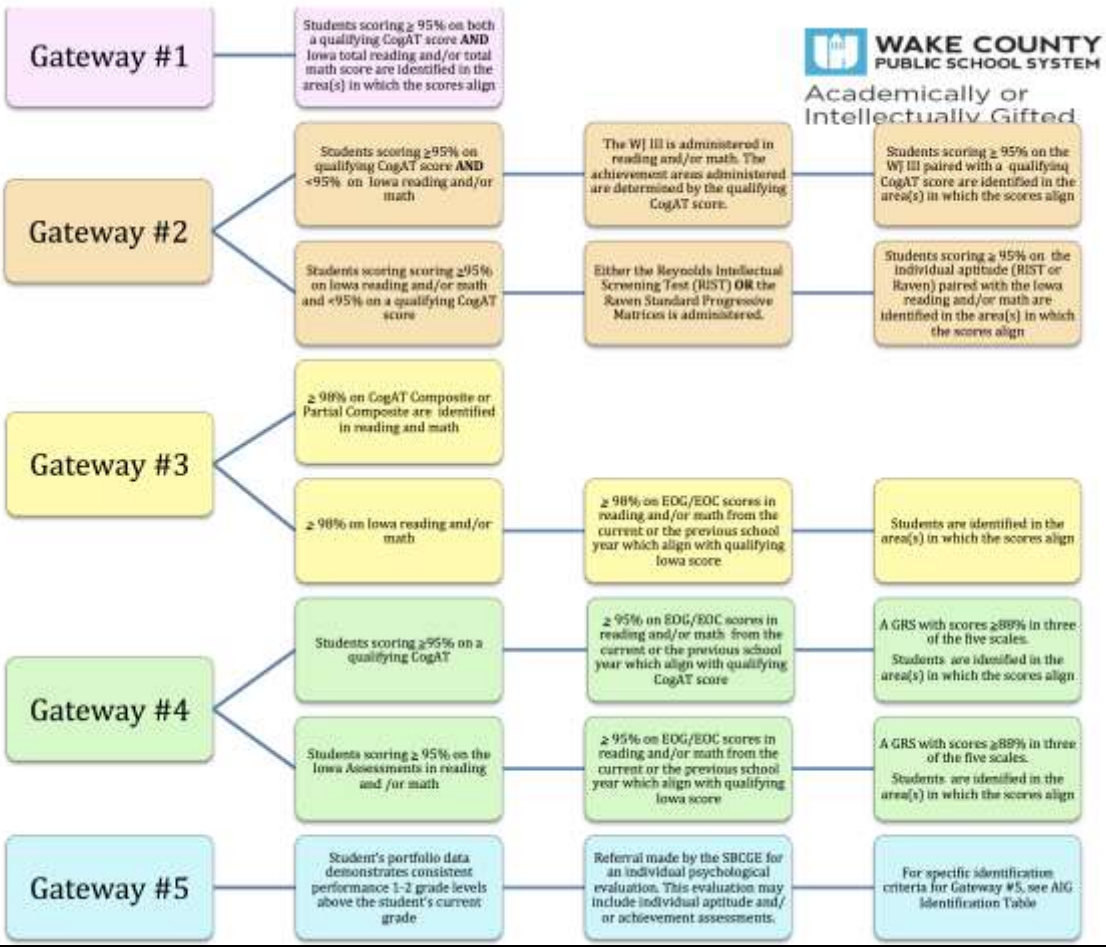
AIG Program Overview

Just over 100,000 students in the district are enrolled in grades that include Academically or Intellectually Gifted (AIG) programming and nearly a quarter of those students are enrolled in AIG programming for mathematics, English/Language Arts (ELA), or both subjects. Gifted education in WCPSS is governed by a statute the North Carolina General Assembly approved in 1996 requiring local education agencies (LEA) across the state to submit three-year AIG plans beginning in the 1998-99 school year. The district is currently in the midst of its eighth plan. More recent plans articulated how the district implements the North Carolina AIG Program Standards, which were adopted in 2009. In particular, Standard 2, "Differentiated Curriculum and Instruction," calls on LEAs to employ "challenging, rigorous, and relevant curriculum... to accommodate a range of academic, intellectual, social, and emotional needs of gifted learners."

The business-as-usual condition for cultivating AIG skills in WCPSS in grades K-2 consists of two primary resources. Primary Education Thinking Skills (PETS) (Nichols, 1997), a differentiated learning approach based in Bloom's Taxonomy, and the district-developed Kids into Thinking (KIT) modules. For PETS, teachers are expected to complete two whole-class lessons and one small group extension lesson per quarter. KIT-related activities are conducted at the teacher's discretion. While the control condition activities are designed to cultivate gifted potential, they are primarily content-based and do not facilitate a school-wide culture of gifted expectations. Teachers have some discretion about specifically when and how they use these programs and district staff does not systematically monitor implementation fidelity. The Nurturing for a Bright Tomorrow (NBT) curriculum is designed to achieve school-level consistency of gifted practices in the early elementary grades through the intersection of its three components, as well as through teacher training and continued school- and classroom-level implementation monitoring.

In order to successfully screen for AIG status in WCPSS, all entering 3rd graders are administered a universal screener. To qualify, students must first score at or above the 95th percentile on the Cognitive Ability Test (CogAT) and subsequently at or above the 95th percentile on the Iowa Assessments (Iowa). Students who meet this threshold sequence are referred to the School Based Committee for Gifted Education (SBCGE), which approves, denies, or gathers additional data for use in AIG identification decisions. The additional data may include testing using individual aptitude or achievement assessments. At the time NBT was launched, this was the primary "gateway" (Figure 1) through which students could be identified as gifted. In the years since, the district has expanded its number of gateways to include four additional paths to giftedness. Given the shifting dynamics of gifted identification criteria in the district and to simplify our analyses, we report the impact of NBT on overall gifted identification in math, reading, both subjects, and any combination of the three.

Figure 1
WCPSS AIG Gateways 2013-14 to 2015-16



Source: "WCPSS Local AIG Plan, 2013-2016: A Plan for Meeting the Needs of Academically or Intellectually Gifted Students," Appendix C, p. 100.

Nurturing for a Bright Tomorrow Overview

Nurturing for a Bright Tomorrow (NBT) represents the third iteration of what was previously called Project Bright IDEA. The program builds on an earlier partnership between the North Carolina Department of Instruction (NCDPI) and the American Association for Gifted Students (AAGS) at Duke University that resulted in Project Bright IDEA. The motivation behind the earlier versions and the current NBT intervention remain the same: to develop the interests and abilities of traditionally underserved groups in the early elementary grades so that by grade 3, they are equipped to qualify for and excel in gifted education programs. The difference between Project Bright IDEA and Nurturing for a Bright Tomorrow is that the latter was randomly assigned to schools and implemented in a way that could generate causal evidence. While anecdotal and descriptive evidence suggests that the two iterations of Project Bright IDEA (a pilot from 2000-2003 and scale-up from 2004-2009) positively influenced achievement and gifted identification, neither project phase produced causal impact estimates (Watson, Hargett, & Tzur, 2010).

NBT is designed to train early elementary school teachers (K-2) to develop the skills and expectations required to help children to attain gifted identification in elementary school. A large body of evidence supports the belief that variations in teacher disposition toward students can significantly influence perceived or real student outcomes (Anderson-Clark, Green, & Henley, 2008; Dee, 2004; Grissom & Redding, 2016; Love & Kruger, 2005; McKown & Weinstein, 2008; VanTassel-Baska, & Brown, 2007). NBT is designed to influence these dispositions through a comprehensive curricular approach that includes three components: Building Thinking Skills & Key Concepts (Parks & Black, 1997), Habits of Mind (Costa & Kallick, 2009), and Task Rotation (Silver, Jackson, & Moirao, 2011).

According to The Critical Thinking Co., the publisher of Building Thinking Skills, this book series prepares students for gifted testing by teaching them three core processes: (1) observing, recognizing, and describing characteristics, (2) distinguishing similarities and differences, and (3) identifying and completing sequences, classifications, and analogies. At the core of the Habits of Mind framework are 16 attributes students are expected to internalize as they develop gifted behaviors. According to its developer,

> *"When we teach for the Habits of Mind, we are interested also in how students behave when they don't know an answer. The Habits of Mind are performed in response to questions and problems, the answers to which are not immediately known. We are interested in enhancing the ways students produce knowledge rather than how they merely reproduce it. We want students to learn how to develop a critical stance with their work: inquiring, editing, thinking flexibly, and learning from another person's perspective. The critical attribute of intelligent human beings is not only having information but also knowing how to act on it" (Costa, 2009).*

Task Rotation is a teaching strategy in which students cycle through four different styles of activities: mastery tasks, interpersonal tasks, understanding tasks, and self-expressive tasks. In practice, students are given an activity that is differentiated by task style, and students rotate through each task (Silver et al., 2011).

Taken together, the WCPSS-Duke implementation team hypothesized that these integrated approaches in early elementary school classrooms would provide teachers with an overarching framework to differentiate instruction, teach advanced vocabulary and speaking skills, and build sustainable approaches to problem solving. These changes in instruction were then hypothesized to lead to higher levels of AIG identification, particularly among under-represented groups.

Teacher training and implementation

The training for NBT was intended to help teachers create scholarly environments that engage all students in sophisticated investigations of materials and performance tasks, and to understand and apply advanced critical and creative thinking. As a result, the classroom environment was transformed into differentiated problem-solving centers to increase background knowledge for all students so they could meet performance tasks. Kindergarten teachers received training during the summer prior to the first year (2014-15) of implementation, with first and second grade teachers receiving the training during the second and third years of implementation, respectively. Teachers from each grade level received a second training session, which served as a "refresher" during the summer following their first year of program implementation. Teachers received stipends for attending the training sessions. The NBT training sessions changed over the course of the project. Primary changes have been in the quality of the training, how NBT is introduced to teachers, and the amount of support and feedback teachers receive. Initially, external trainers (not affiliated with the district) conducted training during two full-day workshops. However, teachers' feedback suggested that this form of training introduced NBT in a manner that was overwhelming; they were being asked to absorb three rather large components (i.e., Thinking Skills, Task Rotation, and Habits of Mind)—each presented separately. This gave teachers the impression that these components were entirely independent, which could be a major challenge given competing demands on their time.

In response to feedback from teachers, the training was re-framed as one educational model of instruction with three components that teachers integrated when conducting regular classroom lessons. The Duke-WCPSS implementation team modified the training for year 2 (the first kindergarten cohort as first graders) to reflect this new understanding of the intervention. However, external trainers continued to conduct the training, which post-training feedback from teachers suggested was ineffective. As such, WCPSS' AIG teachers conducted the training for years 3 and 4. Additionally, the training was condensed to one day with four one-hour sessions. Although these adjustments improved satisfaction with teacher training, they represented three different sets of experiences with the intervention for teachers. These

differences coincided with grade level, with kindergarten teachers experiencing the most challenges while second grade teachers acclimated more seamlessly to the intervention. The challenges encountered during the first two years of NBT were associated with the installation of a completely new model of instruction in a large school system. However, these challenges yielded strategies that enabled the collaborative team to make adjustments to the training that they could not have anticipated without actual implementation.

Data and Analysis Plan

NBT was randomly assigned to 16 elementary schools (out of 32 total schools) on the basis of gifted identification rates in math and reading. Schools with the largest gaps between expected and actual identification rates for Black and Hispanic/Latino students compared with their White counterparts were initially deemed eligible for NBT. However, these schools already had high levels of gifted identification for White students and implementing NBT at these schools had the potential to further help raise gifted rates at the expense of schools with low rates for all students. We therefore selected the 52 schools with gifted rates below the district median rate of 4% (in both math and reading) regardless of rates by race/ethnicity. We reached out to these 52 schools and selected 32 schools that committed to implementing the program with fidelity if they received it through random assignment. We then sorted schools on the basis of each school's gifted identification rate and randomly assigned NBT within school pairs.

Table 2 shows pre-NBT balance between the treatment and control groups on a range of characteristics, including student-level indicators and relevant school-level characteristics. The treatment and control groups were similar, on average, across most measures. The notable exception was the percentage of students in each group attending year-round schools. That percentage was significantly higher for students in the control group—roughly two-thirds compared to nearly 30 percent in the treatment group. There is no strong theoretical reason to suggest that NBT would be more effective in schools with one type of calendar or another, so we simply controlled for this variation in enrollment by calendar type in the impact models. The percentage of males enrolled in each group varied slightly and, similar to the differences by calendar, we are not overly concerned by these differences and simply control for them in the statistical models.

Table 2
Pre-NBT Balance between Treatment and Control Groups, Student and School Characteristics

Variable	(1) Control	(2) Treatment	(3) Difference (T-C)	(4) P-value
Student Characteristics				
Male	0.506	0.524	0.017	0.083*
Asian	0.040	0.040	0.000	0.983
Hispanic/Latino	0.219	0.270	0.050	0.180
Black	0.326	0.340	0.014	0.814
White	0.380	0.316	-0.065	0.439
Other Race	0.035	0.034	-0.000	0.952
LEP	0.135	0.159	0.024	0.448
SWD	0.082	0.085	0.003	0.787
School Characteristics				
Magnet School	0.194	0.152	-0.042	0.744
Year-Round Calendar	0.662	0.284	-0.378	0.032**
Title I	0.524	0.670	0.147	0.425
School SES	0.373	0.404	0.030	0.569
Students	3,537	3,484	7,021	
Schools	16	16	32	

Notes: This table reports student-level and relevant school-level means for students in NBT and control schools at the beginning of the 2014-15 (Cohort 1) and 2016-17 (Cohort 2) school years. Column (1) displays means for the control group, Column (2) displays means for the treatment group, and Column (3) displays the difference between the two. Column (4) reports associated p-values. Standard errors are clustered at the school level. * p<0.10 ** p<0.05 *** p<0.01

Table 3 shows standardized achievement differences between the treatment and control groups at the beginning of NBT's launch. In the overall sample of test scores that includes all students in the district, the average is set to zero and standard deviation is set to one. Thus, Columns (1) and (2) show the standardized difference in DIBELS scores between the control and NBT groups relative to the district average. These results show that both the treatment and control groups scored roughly 0.20 standard deviations (SD) lower than district averages on DIBELS assessments. This is a moderate difference and an expected one, since NBT was randomly assigned among schools with low gifted identification rates—which correlates strongly with overall student achievement.

Table 3
Pre-NBT Balance between Treatment and Control Groups, DIBELS Performance

Variable	(1) Control	(2) Treatment	(3) Difference	(4) P-value
Achievement Measure				
DIBELS Composite	-0.202	-0.232	-0.030	0.753
DIBELS First Sound Fluency (FSF)	-0.207	-0.189	0.018	0.844
DIBELS Letter Naming Fluency (LNF)	-0.166	-0.223	-0.057	0.513
Students	3,423	3,413	6,836	
Schools	16	16	32	

Notes: The results above show the standardized difference in test scores between the control group and the district average (1) and the treatment group and the district average (2). The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are administered to all district students in grades K-3. DIBELS consists of six individual measures and one composite score that provides a weighted average of available tests by grade and period (beginning, middle, or end of the year). Only FSF and LNF are administered at the beginning of kindergarten. As such, the standardized means for these two tests and the associated composite score are presented here. Standard errors are clustered at the school level. * p<0.10 ** p<0.05 *** p<0.01

The business-as-usual condition at the 16 control schools, summarized earlier, includes Primary Education Thinking Skills (PETS) (Nichols, 1997), a differentiated learning approach based in Bloom's Taxonomy, and the district-developed Kids into Thinking (KIT) modules. Our outcomes of interest used to compare the treatment and control schools are the Naglieri Nonverbal Ability Test (NNAT) and gifted identification rates based on CogAT and Iowa thresholds. We present impacts for the two NBT cohorts as they move through their early elementary school years. Table 4 summarizes the movement of cohorts over each cohort's three-year implementation period. Since AIG screening is generally held at the beginning of the school year, we do not consider the 3rd grade year an implementation year for NBT.

Table 4
Movement of NBT Cohorts over Grade Levels and School Years

	2014-15	**2015-16**	**2016-17**	**2017-18**	**2018-19**	**Key outcomes**
Kindergarten	Cohort 1	Cohort 2				None
1st Grade		Cohort 1	Cohort 2			NNAT
2nd Grade			Cohort 1	Cohort 2		NNAT
3rd Grade				Cohort 1	Cohort 2	AIG

Students were administered the 2nd Edition of the NNAT at the end of grade 1 and a re-normed, 3rd Edition of the NNAT at the end of grade 2. The NNAT is a nonverbal ability measure designed to predict achievement for students across a diverse range of backgrounds. The test assesses student competencies across four domains: Pattern Completion, Reasoning by Analogy, Serial Reasoning, and Spatial Visualization. We used NNAT as an interim measure of gifted potential prior to formal identification for the two cohorts in grade 3. Studies of the first edition of NNAT suggest that the test strongly predicted gifted identification for diverse populations (Lidz & Macrine, 2001; Naglieri & Ford, 2003) and performed similarly to the CogAT6 achievement test (Giessman et al., 2013). However, some researchers have cautioned against the use of NNAT as the sole criteria for identifying gifted populations (Giessman et al., 2013; Carman & Taylor, 2009; Lakin, 2012; Lakin & Lohman, 2011; Lohman, 2005). Such caution has led us to utilize NNAT only as secondary measure to inform the potential effectiveness of NBT prior to formal identification.

In WCPSS, the CogAT is administered to all 3rd graders in the fall, and students who meet the 95th percentile cutoff are then eligible to take the Iowa Assessment. The district's 3-year AIG plan includes a complex series of "gateways" designed to increase the proportion of traditionally under-identified students. Rather than test all the gateways—which we cannot do given the absence of certain secondary and tertiary measures—we construct four relevant outcomes for students ultimately identified as gifted through any gateway.

Our sample consists of roughly 7,000 students in two cohorts enrolled in 32 schools that were randomly assigned to treatment or control groups. We examined the impact of NBT on the NNAT in three ways: (1) on the NNAT scale score, (2) on the rates of gifted identification based on students meeting the 95th and 98th percentiles on the NNAT (2nd and 3rd editions), and (3) on the rates of gifted identification based on students meeting any AIG gateway threshold.

Results

The impacts of Nurturing for a Bright Tomorrow were mixed. The program caused NNAT scores to rise, but it did not meaningfully impact the likelihood that students enrolled in NBT schools would be identified as AIG. This was true for informal gifted thresholds established through the NNAT as well as formal thresholds established through the CogAT and Iowa tests. There were a few exceptions for certain subgroup classifications, but as exceptions, they did not change the overall story. This section continues with a summary of the key findings. The Appendix includes a brief technical discussion of how we measured impacts.

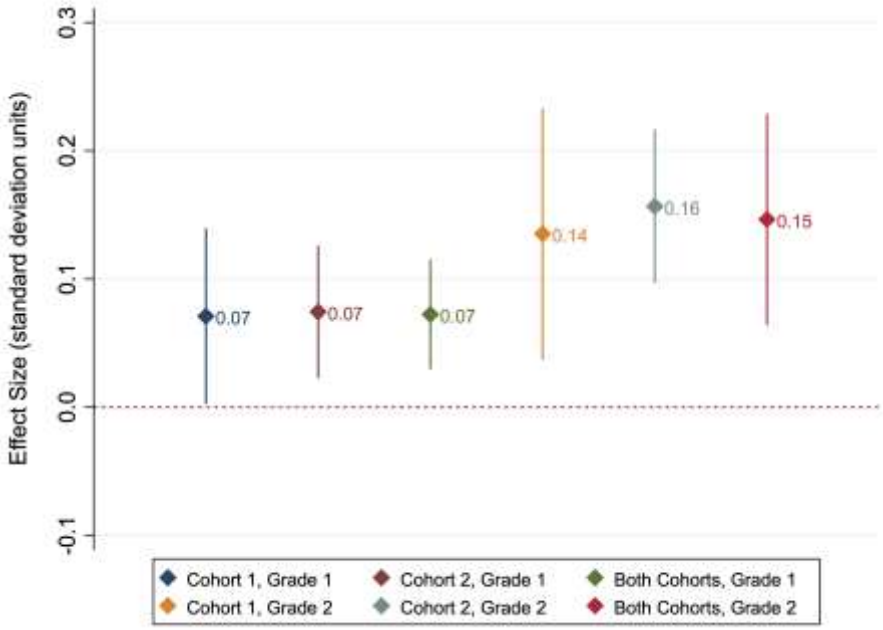
NNAT Scores

Figure 2 shows the impact of NBT on NNAT test scores. The first three markers show that among students initially enrolled in an NBT school as kindergarten students, NBT caused their NNAT scores to increase by 0.07 standard deviations (*SD*). This is true for 1st graders in both

cohorts. An effect of 0.07 *SD* is small and represents an increase of roughly 3 percentile points from the mean. For comparison, the well-known Tennessee STAR class size intervention caused test scores to increase by about 0.20 *SD*. For further comparison, an analysis of hundreds of broad curricular interventions and dozens of schoolwide interventions revealed an effect of roughly 0.10-0.14 *SD* (Lipsey et al., 2012), or about 6 percentile points higher than the mean. Thus, while the magnitude of the effect for 1st graders was relatively small, the size of the effect for 2nd graders exceeded benchmark levels. The next set of markers in Figure 2 shows that the effect of NBT on grade 2 NNAT scores ranged from 0.14-0.16 *SD*. The effect for 2nd graders may be smaller than that of Tennessee STAR, for example, but larger than the average effect size drawn from hundreds of comparable interventions.

Hispanic/Latino students (Appendix Figure A1) in the treatment group outperformed their counterparts by an even greater magnitude than the overall sample, with effects ranging from 0.15 for 1st graders in Cohort 1 to 0.44 *SD* for second graders in Cohort 2. Black students in the treatment group performed similarly compared to their control group counterparts.

Figure 2
Impact of Nurturing for a Bright Tomorrow on Naglieri Nonverbal Ability Test (NNAT) Scores

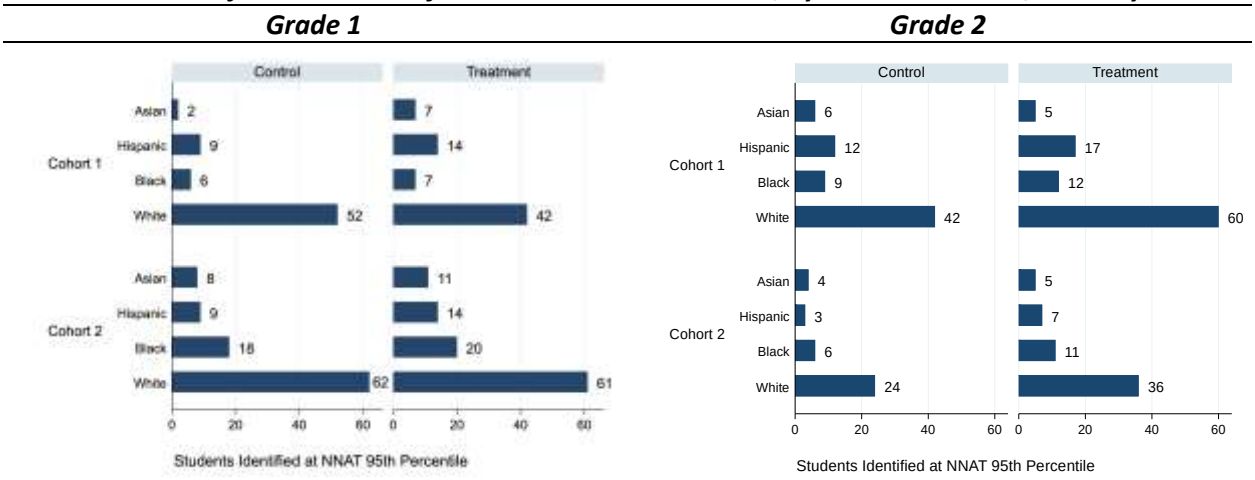


Note: Results are for students in each of the two kindergarten cohorts present at the beginning of the 2014-15 and 2015-16 school years when NBT was launched. Markers represent the effect of Nurturing compared to the control group. Effect sizes above and below the dotted red are statistically significant at the 5% level if the vertical 95% confidence interval (CI) bands do not touch the dotted red line. If the 95% CI bands do touch the red line, the effect is not statistically different from zero. The NNAT was administered to students in each cohort in the spring of grades 1 and 2.

NNAT Gifted Identification Counts and Odds

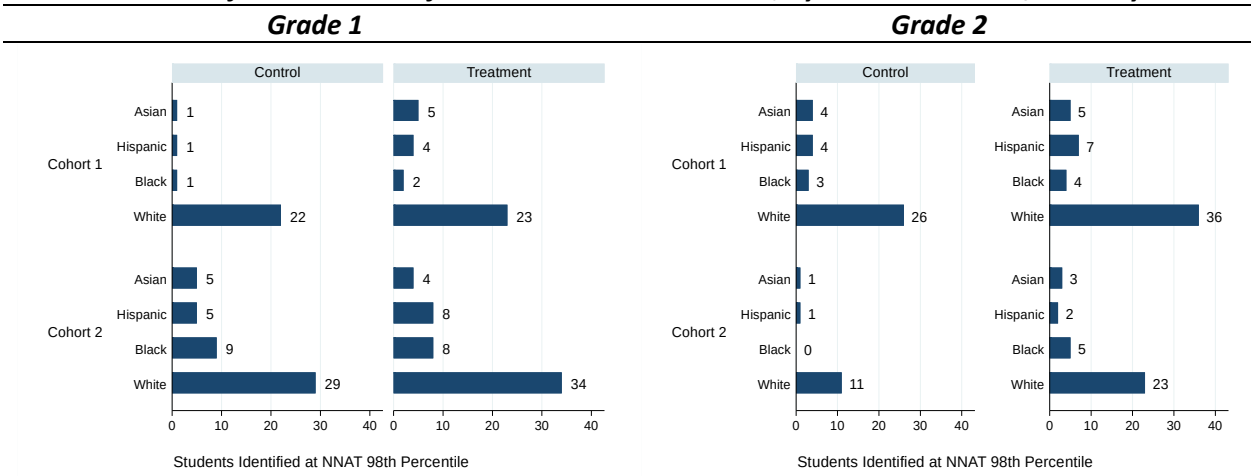
In addition to measuring the effect of NBT on NNAT test scores, we also tabulated the number of students across racial/ethnic groups who qualified as gifted based on NNAT thresholds. Figures 3, 4, 6, and 7 display counts and Appendix Table A1 presents these counts represented as rates. We did this for both the 95th percentile and the 98th percentile. The left side of Figure 3 shows that by grade 1, the number total of NBT ("Treatment") students in both cohorts identified as gifted (176) slightly exceeded the number in the control group (166). While the largest number of students identified across both the treatment and control group was White, the treatment group did have a larger number of Black and Hispanic/Latino students in each cohort who met the NNAT's gifted thresholds. The right side of Figure 3 shows that among 2nd graders, more students in the treatment group met the NNAT gifted threshold and in each cohort, more Black and Hispanic/Latino students were identified in the treatment group compared with the control group. Figure 4 shows that these results were largely consistent at the 98th percentile.

Figure 3
Number of Students Identified at NNAT 95th Percentile, by Cohort and Race/Ethnicity



Note: The bars above represent the number of kindergarten students in the treatment and control groups scoring at or above the 95th percentile when they took the NNAT.

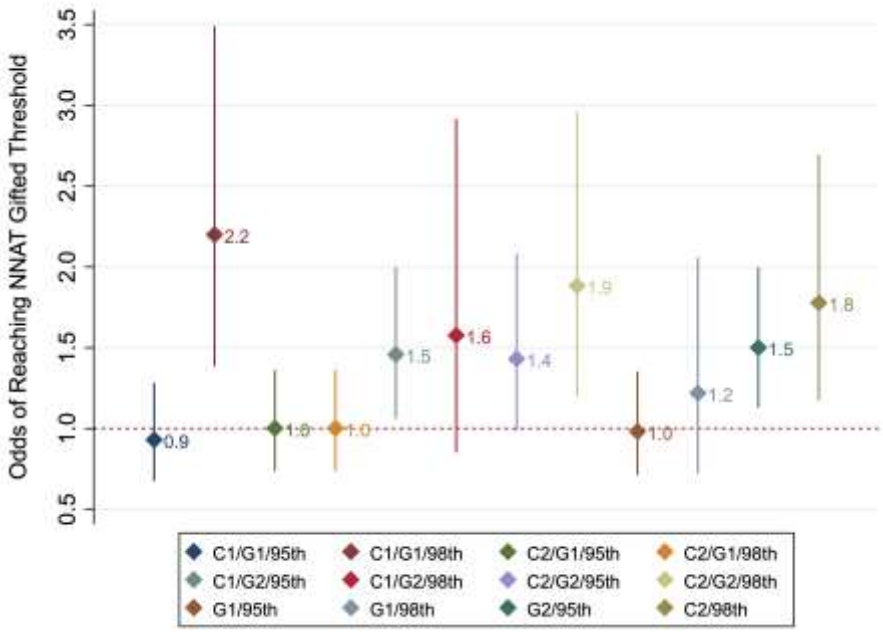
Figure 4
Number of Students Identified at NNAT 98th Percentile, by Cohort and Race/Ethnicity



Note: The bars above represent the number of kindergarten students in the treatment and control groups scoring at or above the 98th percentile when they took the NNAT.

In order to measure whether the differences in the number of students identified at the NNAT thresholds between the treatment and control groups were statistically significant, we estimated the odds of identification. To illustrate, an odds ratio (OR) of 1.0 indicates no difference between the treatment and control groups. An odds ratio of 0.5 suggests that the treatment group had a 50% lower odds of reaching the gifted threshold, while an odds ratio of 1.5 means the treatment group had a 50% greater odds of reaching the gifted threshold. Figure 5 shows that the odds of most cohort-grade treatment groups reaching the gifted threshold exceeded 1.0. In the legend, "C1/G1/95th," for example, refers to the 1st cohort of 1st graders who reached the 95th percentile on the NNAT. In particular, 1st (OR = 2.2) and 2nd graders (OR = 1.5) from Cohort 1 each had a statistically significant greater odds of reaching the 95th percentile on the NNAT than students in the control group. Overall, 2nd graders had a significant odds of reaching the 95th (OR = 1.5) and 98th (OR = 1.8) percentiles. Due to the small number of students from subgroups identified in Figures 3-4, we could not precisely estimate odds ratios for subgroups.

Figure 5
Impact of Nurturing for a Bright Tomorrow on
Odds of Reaching NNAT Gifted Identification Thresholds



Note: Results are for students in each of the two kindergarten cohorts present at the beginning of the 2014-15 and 2015-16 school years when NBT was launched. Markers represent the effect of NBT compared to the control group. Odds ratios above and below the dotted red line are statistically significant at the 5% level if the vertical 95% confidence interval (CI) bands do not touch the dotted red line. If the 95% CI bands do touch the red line, the effect is not statistically different from zero. "C" = Cohort, "G" = Grade, and 95th and 98th represent NNAT percentiles.

AIG Gifted Identification Counts and Odds

The primary outcome of interest is whether more students enrolled in NBT schools were identified as gifted based on the district's formal qualifying criteria. Recall that since the 2013-2016 AIG plan was adopted, the district added an additional gateway to gifted identification. The results in this section reflect AIG classification through any gateway. Here, we report four separate outcomes: (1) identification in math, (2) identification in reading, (3) identification in both math and reading, and (4) identification in any combination of subjects (i.e., math, reading, or both). The following discussion will reference Figures 6-8, which are interpreted similarly to NNAT Figures 3-5, above.

Table 5 shows the descriptive results by cohort and treatment condition for each of these four AIG classification methods. These results are for reference and do not represent the causal effects of NBT, which are displayed in Figure 8 and discussed below. Across most treatment-cohort combinations, the control group ended with higher gifted identification rates than the treatment group, though rarely did these differences reach conventional levels of statistical significance.

Table 5
Gifted Identification Rates (%), by Treatment Condition and Cohort

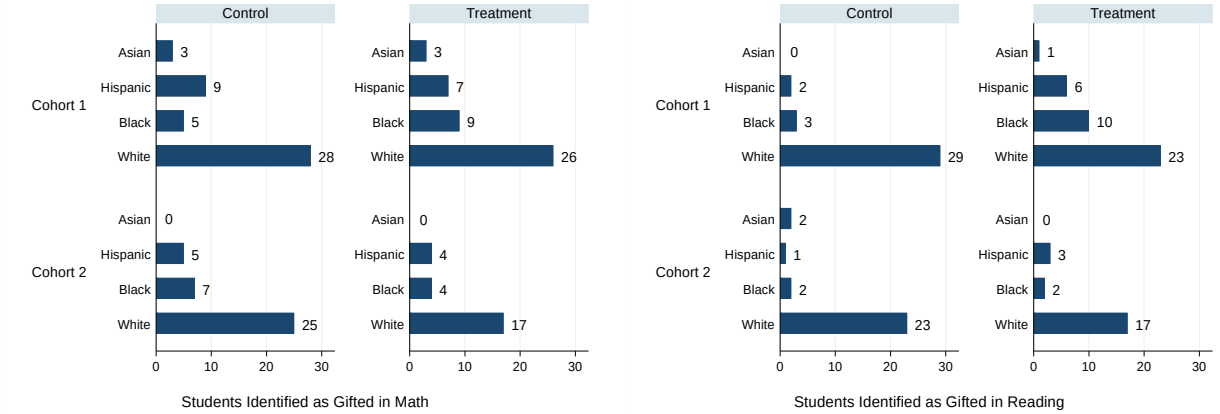
	AIG Math	AIG Reading	AIG Both	AIG Any
Treatment	2.1	1.8	3.0	6.9
Cohort 1	2.6	2.3*	3.2	8.1
Cohort 2	1.5	1.3	2.8	5.6
Control	2.5	1.8	4.2*	8.6
Cohort 1	2.7	2.1	4.7*	9.5
Cohort 2	2.2	1.7	3.8	7.7

Notes: these percentage rates apply to roughly 3,500 students across two cohorts in each of the treatment and control groups. The * indicates that the difference in rates between the treatment and control group is statistically significant in terms of odds ratios, which are reported in Figure 8 of this report. All other differences between the treatment and control group within each cohort do not statistically vary from zero.

Beginning with AIG Math, the left side of Figure 6 shows that across both cohorts, 70 students in the treatment group were identified compared with 82 students in the control group. This difference is the opposite of what we would expect from the NBT intervention. Figure 8 shows that while the odds of treatment group students being identified as gifted in math is roughly 10-30% lower than in the control group, this difference is not statistically different from zero, since the marker of confidence intervals intersect with the "null" line of OR = 1.0. Across both cohorts, 62 students were identified as gifted in reading in each of the experimental conditions, treatment and control. However, the odds of students from Cohort 1 were significantly higher (OR = 1.7) because more treatment group students were identified (40 vs. 34). Students in the treatment group had odds roughly 30% lower of being identified as gifted in both reading and math (OR = 0.7) compared with their control group counterparts. This result was driven by statistically significantly lower odds for Cohort 1. Finally, the odds of being identified across all three groups was lower for students in the treatment group (OR = 0.8-0.9), but this difference with the control group was not statistically different from zero.

Figure 6
Number of Students Identified as Gifted, by Cohort and Race/Ethnicity

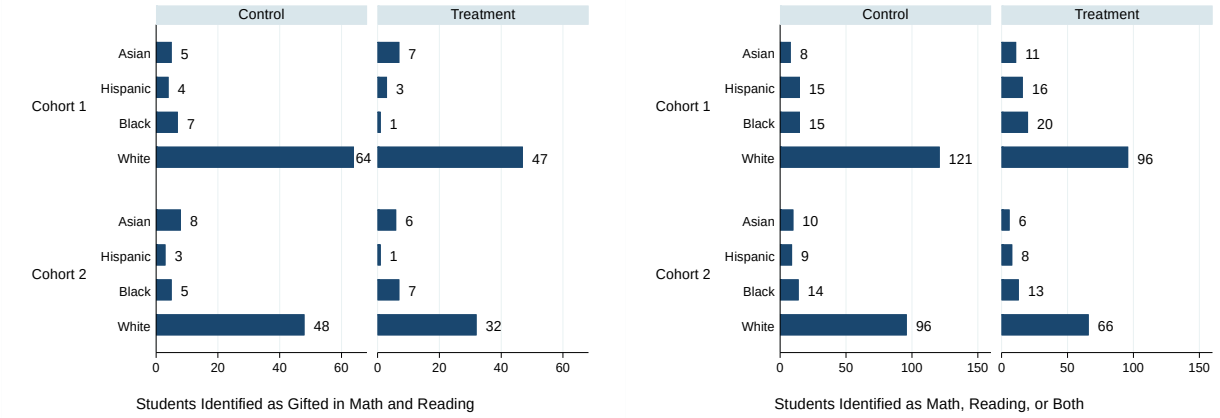
AIG Math ***AIG Reading***



Note: The bars above represent the number of kindergarten students in the treatment and control groups qualifying for gifted services in math (left) or reading (right).

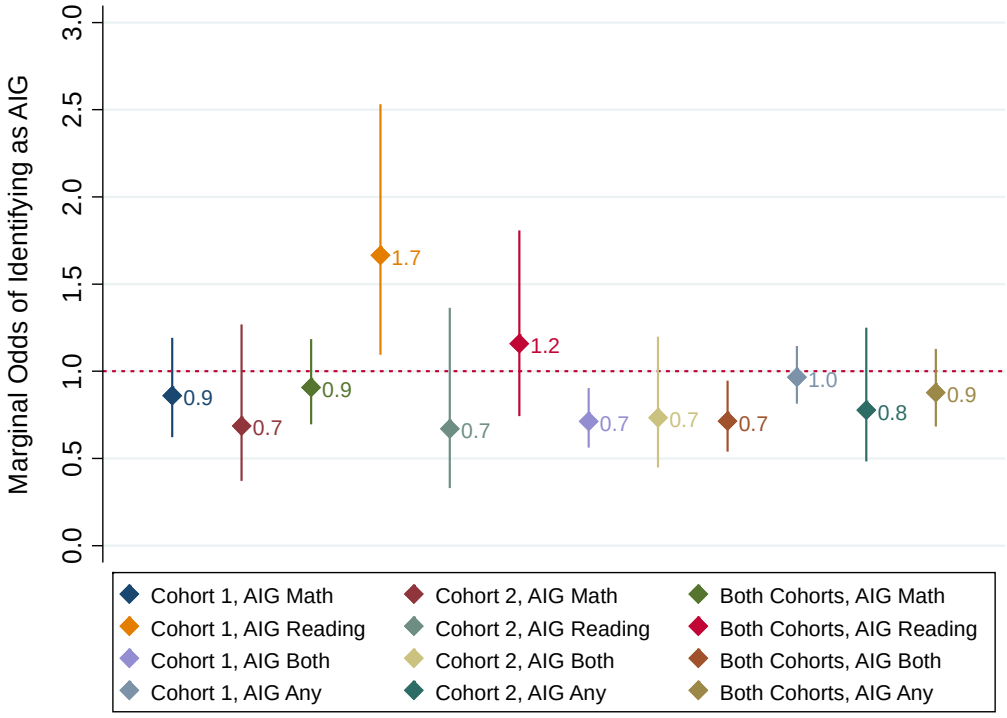
Figure 7
Number of Students Identified as Gifted, by Cohort and Race/Ethnicity

AIG Both ***AIG Any***



Note: "AIG Both" refers to students who identified as gifted in both reading and math. "AIG Any" refers to students who identified as gifted in either math, reading, or both subjects. The bars above represent the number of kindergarten students in the treatment and control groups qualifying for gifted services in both (left) or any (right).

Figure 8
Impact of Nurturing for a Bright Tomorrow on Likelihood of Reaching AIG Gifted Status



Note: Results are for students in each of the two kindergarten cohorts present at the beginning of the 2014-15 and 2015-16 school years when NBT was launched. Markers represent the effect of NBT compared to the control group. Odds ratios above and below the dotted red line are statistically significant at the 5% level if the vertical 95% confidence interval (CI) bands do not touch the dotted red line. If the 95% CI bands do touch the red line, the effect is not statistically different from zero. "AIG Both" refers to students who were identified as gifted in both reading and math and "AIG Any" refers to students who were identified as gifted in either math, reading, or both subjects.

Conclusion and Recommendations

In response to chronically low gifted identification rates in the majority of elementary schools, especially among Black and Hispanic/Latino students, the Wake County Public School System (WCPSS) launched Nurturing for a Bright Tomorrow (NBT) in Fall 2014 in partnership with Duke University. NBT was a multi-component curricular enhancement program designed to train early elementary school teachers (K-2) to develop the skills and expectations required to help children attain gifted identification by 3rd grade. To implement NBT in pursuit of this goal, a joint implementation team consisting of WCPSS' Academically or Intellectually Gifted (AIG) department and staff from Duke University partnered to train teachers, distribute curricular materials, and monitor implementation. NBT, through random assignment, was launched in 16 elementary schools for two cohorts of participants in Fall 2014 and Fall 2015. The 16 treatment schools were randomly identified from a group of 32 schools that agreed to implement NBT if

selected. The 32 elementary schools were drawn from a larger sample of 52 schools that had overall gifted identification rates below the district average.

We answer three questions in this report: (1) Did NBT cause test scores to increase on the Naglieri Nonverbal Ability Test (NNAT), an intermediate measure of gifted abilities? (2) Did NBT increase the odds of students meeting the NNAT's gifted identification thresholds? (3) Did NBT increase the odds of students meeting the district's formal gifted identification thresholds?

NBT did significantly increase students' NNAT scores by roughly 0.07-0.14 standard deviations (*SD*), an effect that is considered small to moderate and represents an increase of roughly 3-6 percentile points higher than average. The effect was even larger for Hispanic/Latino students. Correspondingly, NBT increased the odds that students would reach gifted identification thresholds on the same NNAT assessment for many cohort-grade combinations. Unfortunately, NBT did not lead to widespread increases in the numbers of students who qualified for the AIG program or the odds that students enrolled in NBT schools would qualify for gifted status. The only cohort-subject group among 12 with higher odds of being identified as gifted was Cohort 1 in reading. The odds of becoming gifted in both reading and math were generally lower for students in NBT schools and across nine other cohort-subject combinations, the odds were similar between NBT schools and non-NBT schools.

These findings highlight the challenges associated with cultivating gifted skills and behaviors for students in schools where gifted identification has traditionally been low. The challenge is particularly acute for Black and Hispanic/Latino students. The results here show that while slightly more Black and Hispanic/Latino students in NBT schools were identified through the district's formal gifted identification process by grade 3, the numbers of qualifiers remain small compared with White students (Figures 3-4, 6-7).

Beginning in the 2018-19 school year, WCPSS discontinued NBT but offered components of the full intervention at select grade levels and classrooms. Still, we learned important lessons about launching a schoolwide, multi-faceted randomized controlled trial across grades K-2 and offer the following recommendations to inform future efforts designed to increase gifted enrollment and close gifted identification gaps:

Identify school-level enrichment and nurturing plans. The status quo approach to enrichment and nurturing has been to use the off-the-shelf Primary Education Thinking Skills (PETS) approach and the district-developed modules. A Spring 2018 survey that included responses from 68 AIG teachers revealed that roughly half used these resources in some or all early elementary grades, suggesting that there are many more approaches being implemented in schools that did not respond to the survey or responded that they do not use them. Moreover, post-NBT, schools were offered the opportunity to use certain components of the original intervention. Therefore, there exists a wide amount of variation in which programs and strategies schools use to cultivate gifted behaviors and skills. The AIG department should create

a comprehensive inventory of current programs in order to devise a post-NBT strategy for closing gifted identification gaps across schools and subgroups.

Consider using the Naglieri Nonverbal Ability Test (NNAT) as a component of identification. Students in NBT schools significantly outperformed their non-NBT counterparts on the NNAT. While the NNAT should not be used exclusively as a gifted qualification assessment, it can be used in combination with traditional measures, notably the CogAT and Iowa assessments. Indeed, the NNAT is already classified as a "non-traditional measure" in the district's 2016-2019 Local AIG Plan. The results from this evaluation suggest that Hispanic/Latino students, in particular, demonstrated gifted skills and abilities on the NNAT in ways that they do not on traditional measures such as the CogAT and Iowa Assessments. Emphasizing its relative use could help narrow the gifted identification gap among White and Hispanic/Latino students, in particular.

References

Anderson-Clark, T. N., Green, R. J., & Henley, T. B. (2008). The relationship between first names and teacher expectations for achievement motivation. *Journal of Language and Social Psychology*, *27*(1), 94–99.

Carman, C. A., & Taylor, D. K. (2009). Socioeconomic status effects on using the Naglieri Non-Verbal Ability Test (NNAT) to identify the gifted/talented. *Gifted Child Quarterly*.

Costa, A. L. (2009). Describing the habits of mind. In Costa, A. L. and Kallick, B., editors, *Learning and Leading with Habits of Mind: 16 Essential Characteristics for Success*, chapter 2, pages 15–41. Association for Supervision & Curriculum Development.

Costa, A. L., & Kallick, B. (Eds.). (2009). *Learning and leading with habits of mind: 16 essential characteristics for success*. Association for Supervision & Curriculum Development.

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, *86*(1), 195–210.

Giessman, J. A., Gambrell, J. L., & Stebbins, M. S. (2013). Minority performance on the Naglieri Nonverbal Ability Test, versus the Cognitive Abilities Test, form 6 one gifted program's experience. *Gifted Child Quarterly*, *57*(2), 101–109.

Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality. *AERA Open*, *2*(1).

Lakin, J. M. (2012). Assessing the cognitive abilities of culturally and linguistically diverse students: Predictive validity of verbal, quantitative, and nonverbal tests. *Psychology in the Schools*, *49*(8), 756–768.

Lakin, J. M., & Lohman, D. F. (2011). The predictive accuracy of verbal, quantitative, and nonverbal reasoning tests: Consequences for talent identification and program diversity. *Journal for the Education of the Gifted*, *34*(4), 595–623.

Lidz, C. S., & Macrine, S. L. (2001). An alternative approach to the identification of gifted culturally and linguistically diverse learners the contribution of dynamic assessment. *School Psychology International, 22*(1), 74–96.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., … Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. National Center for Special Education Research.

List, J. A., Sadoff, S., & Wagner, M. (2011). So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics, 14*(4), 439–457.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly*, *49*(2), 111–138.

Love, A., & Kruger, A. C. (2005). Teacher beliefs and student achievement in urban schools serving African American students. *The Journal of Educational Research*, *99*(2), 87–98.

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of school psychology*, *46*(3), 235–61.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly*, *47*(2), 155–160.

Nichols, J. (1997). *Primary Education Thinking Skills #3*. Marion, IL: Pieces of Learning.

Parks, S., & Black, H. (1997). *Building thinking skills*. North Bend, OR: The Critical Thinking Co.

Silver, H. F., Jackson, J. W., & Moirao, D. R. (2011). *Task rotation: Strategies for differentiating activities and assessments by learning style*. ASCD.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse.. (2014). WWC procedures and standards handbook, version 3.0.

VanTassel-Baska, J., & Brown, E. F. (2007). Toward best practice: An analysis of the efficacy of curriculum models in gifted education. *Gifted Child Quarterly*, *51*(4), 342–358.

Watson, M., Hargett, M. P., & Tzur, R. (2010). Project Bright Idea 2: Interest Development Early Abilities. Raleigh, NC: North Carolina Department of Public Instruction.

Appendix

To measure the impact of Nurturing for a Bright Tomorrow (NBT), we fit regression models that model NNAT test scores and the odds of being identified as a function of NBT, controlling for student-level characteristics and matched-pair fixed effects. Because NBT was assigned at the school level, we cluster standard errors at the school level, which accounts for variation across schools. The model helps us answer the question, What would happen if all WCPSS elementary schools with relatively low gifted identification rates were offered NBT?

To estimate the impact of NBT on NNAT test scores, we fit the following model:

$$NNAT_{ij} = \beta_0 + \beta_1 NBT_{ij} + Z_{ij} + \varepsilon_{ij}$$

where $NNAT_{ij}$ represents our dependent variable of interest—i.e., the standardized score on the NNAT—for student *i* in school *j*; β_0 represents the constant; β_1 represents the effect of the treatment (the offer of NBT) for student *i* in school *j*; *Z* represents student- and school-level control variables. The final term represents a student-level error term, representing unobserved variation due to factors we cannot measure. The outcome is expressed in terms of effect sizes.

To estimate the impact of NBT on the odds of being identified as gifted using either the NNAT or AIG program (i.e., CogAT and Iowa), we use a similar setup to measure the impact of NBT on rates of gifted identification based on NNAT thresholds and AIG gateways. However, instead of using linear regression, we use logistic regression since the outcome of interest (gifted status) is either 1 (gifted) or 0 (not gifted).

To estimate the impact of NBT on the odds of meeting gifted thresholds, we fit the following model:

$$ln(\pi/(1-\pi))_{ij} = \beta_0 + \beta_1 NBT_{ij} + Z_{ij} + \varepsilon_{ij}$$

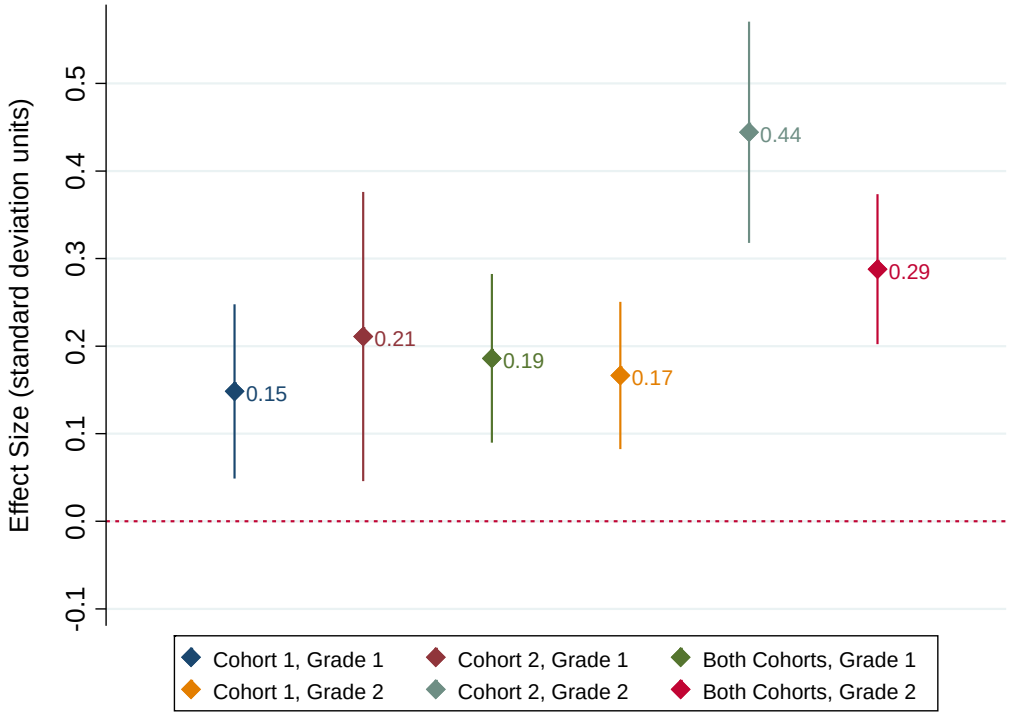
where π represents the probability of attaining gifted status. The above equation expresses the natural logarithm of the odds ("log odds"), which are then transformed into a more easily interpretable odds ratio (OR). If the result is OR > 1, then the odds of NBT students meeting gifted thresholds is higher than the odds of control group students meeting the threshold. If OR = 1, the odds of both events is equal. If OR < 1, then the odds of students in the treatment group meeting gifted thresholds is lower than that of the control group.

Appendix Table A1
Gifted Identification Rates (%) by Condition, Cohort, Benchmark, and Race/Ethnicity

Condition	Cohort	Benchmark	Grade	Asian	Hispanic	Black	White
Control	1	NNAT 95th	1	4.6	3.0	1.4	9.7
Control	2	NNAT 95th	1	24.2	3.3	4.8	13.9
Treatment	1	NNAT 95th	1	13.2	3.7	1.6	9.7
Treatment	2	NNAT 95th	1	22.0	4.8	5.9	17.3
Control	1	NNAT 95th	2	16.2	4.7	2.6	9.4
Control	2	NNAT 95th	2	16.0	1.4	2.1	6.0
Treatment	1	NNAT 95th	2	11.1	4.8	3.4	16.7
Treatment	2	NNAT 95th	2	14.7	2.7	4.1	11.0
Control	1	NNAT 98th	1	2.3	0.3	0.2	4.1
Control	2	NNAT 98th	1	15.2	1.8	2.4	6.5
Treatment	1	NNAT 98th	1	9.4	1.1	0.5	5.3
Treatment	2	NNAT 98th	1	8.0	2.7	2.3	9.6
Control	1	NNAT 98th	2	10.8	1.6	0.9	5.8
Control	2	NNAT 98th	2	4.0	0.5	0.0	2.7
Treatment	1	NNAT 98th	2	11.1	2.0	1.1	10.0
Treatment	2	NNAT 98th	2	8.8	0.8	1.9	7.1
Control	1	AIG Math	3	4.5	2.4	0.9	4.2
Control	2	AIG Math	3	0.0	1.3	1.2	3.7
Treatment	1	AIG Math	3	4.4	1.4	1.5	4.5
Treatment	2	AIG Math	3	0.0	0.9	0.7	3.2
Control	1	AIG Reading	3	0.0	0.5	0.5	4.3
Control	2	AIG Reading	3	2.7	0.3	0.4	3.4
Treatment	1	AIG Reading	3	1.5	1.2	1.6	4.0
Treatment	2	AIG Reading	3	0.0	0.7	0.4	3.2
Control	1	AIG Both	3	7.5	1.1	1.2	9.5
Control	2	AIG Both	3	10.8	0.8	0.9	7.2
Treatment	1	AIG Both	3	10.1	0.6	0.2	8.2
Treatment	2	AIG Both	3	8.5	0.2	1.2	6.1
Control	1	AIG Any	3	11.9	4.0	2.6	17.9
Control	2	AIG Any	3	13.5	2.3	2.4	14.3
Treatment	1	AIG Any	3	15.9	3.1	3.3	16.7
Treatment	2	AIG Any	3	8.5	1.9	2.3	12.6

Note: The rates above correspond to the counts presented in Figures 3, 4, 6, and 7.

Appendix Figure A1
Impact of Nurturing for a Bright Tomorrow on
Naglieri Nonverbal Ability Test (NNAT) Scores, Hispanic Students



Note: Results are for students in each of the two kindergarten cohorts present at the beginning of the 2014-15 and 2015-16 school years when NBT was launched. Markers represent the effect of Nurturing compared to the control group. Effect sizes above and below the dotted red are statistically significant at the 5% level if the vertical 95% confidence interval (CI) bands do not touch the dotted red line. If the 95% CI bands do touch the red line, the effect is not statistically different from zero. The NNAT was administered to students in each cohort in the spring of grades 1 and 2.