EdWorkingPaper No. 22-603

Promises, Pitfalls, and Tradeoffs in Identifying Gifted Learners: Evidence from a Curricular Experiment

Angel H. Harris
Duke University

Darryl V. Hill
Bill & Melinda Gates Foundation

Matthew A. Lenard
Harvard University

Disparities in gifted representation across demographic subgroups represents a large and persistent challenge in U.S. public schools. In this paper, we measure the impacts of a school-wide curricular intervention designed to address such disparities. We implemented Nurturing for a Bright Tomorrow (NBT) as a cluster randomized trial across elementary schools with the low gifted identification rates in one of the nation's largest school systems. NBT did not boost formal gifted identification or math achievement in the early elementary grades. It did increase reading achievement in select cohorts and broadly improved performance on a gifted identification measure that assesses nonverbal abilities distinct from those captured by more commonly used screeners. These impacts were driven by Hispanic and female students. Results suggest that policymakers consider a more diverse battery of qualifying exams to narrow disparity gaps in gifted representation and carefully weigh tradeoffs between universal interventions like NBT and more targeted approaches.

VERSION: July 2022

Suggested citation: Harris, Angel H., Darryl V. Hill, and Matthew A. Lenard. (2022). Promises, Pitfalls, and Tradeoffs in Identifying Gifted Learners: Evidence from a Curricular Experiment. (EdWorkingPaper: 22-603). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/a3pd-6y72

Promises, Pitfalls, and Tradeoffs in Identifying Gifted Learners: Evidence from a Curricular Experiment*

Angel H. Harris[†] Darryl V. Hill[‡] Matthew A. Lenard[§]

July 7, 2022

Disparities in gifted representation across demographic subgroups represents a large and persistent challenge in U.S. public schools. In this paper, we measure the impacts of a school-wide curricular intervention designed to address such disparities. We implemented Nurturing for a Bright Tomorrow (NBT) as a cluster randomized trial across elementary schools with the low gifted identification rates in one of the nation's largest school systems. NBT did not boost formal gifted identification or math achievement in the early elementary grades. It did increase reading achievement in select cohorts and broadly improved performance on a gifted identification measure that assesses non-verbal abilities distinct from those captured by more commonly used screeners. These impacts were driven by Hispanic and female students. Results suggest that policymakers consider a more diverse battery of qualifying exams to narrow disparity gaps in gifted representation and carefully weigh tradeoffs between universal interventions like NBT and more targeted approaches.

*Acknowledgments: This research is supported by the Jacob K. Javits Gifted and Talented Students Education Program (2014; PR/Award # S206A140032) to Duke University and approved by Harvard University's IRB (# 21-0335). The authors thank the project implementation team, which consists of staff from both the Wake County Public School System (WCPSS) and Duke University. They include current and former members of the district's Academically or Intellectually Gifted (AIG) department, notably Shani Brown, Melba Spivey, Alena Treat and staff from Duke's Social Science Research Institute (SSRI), notably Kamilah Legette and Jackie Terrell. Thanks to Meredith Chandler, Margaret Gayle, Christine Lange, Dan Turner, and members of the Duke Bass Connections network for their contributions to training and/or project implementation. Kaitlin Anderson, James Carter III, Mark Chin, Christopher Cleveland, Sandy Darity, Charles Gale, Heather Hill (and seminar participants), Fiona Hollands, Joy Lesnick, Atsuko Muroga, Yilin Pan, Scott Peters, Mikko Silliman, and Ruth Steidinger, along with participants from AEFP, AERA, APPAM, and SREE conference proceedings, provided valuable feedback at various stages of this work. The views expressed herein are those of the authors do not necessarily reflect the views of the Duke University or WCPSS. All errors are ours.

[†] Duke University, angel.harris@duke.edu
[‡] Bill & Melinda Gates Foundation, darryl.hill@gatesfoundation.org
[§] Harvard University, mlenard@g.harvard.edu (corresponding author)

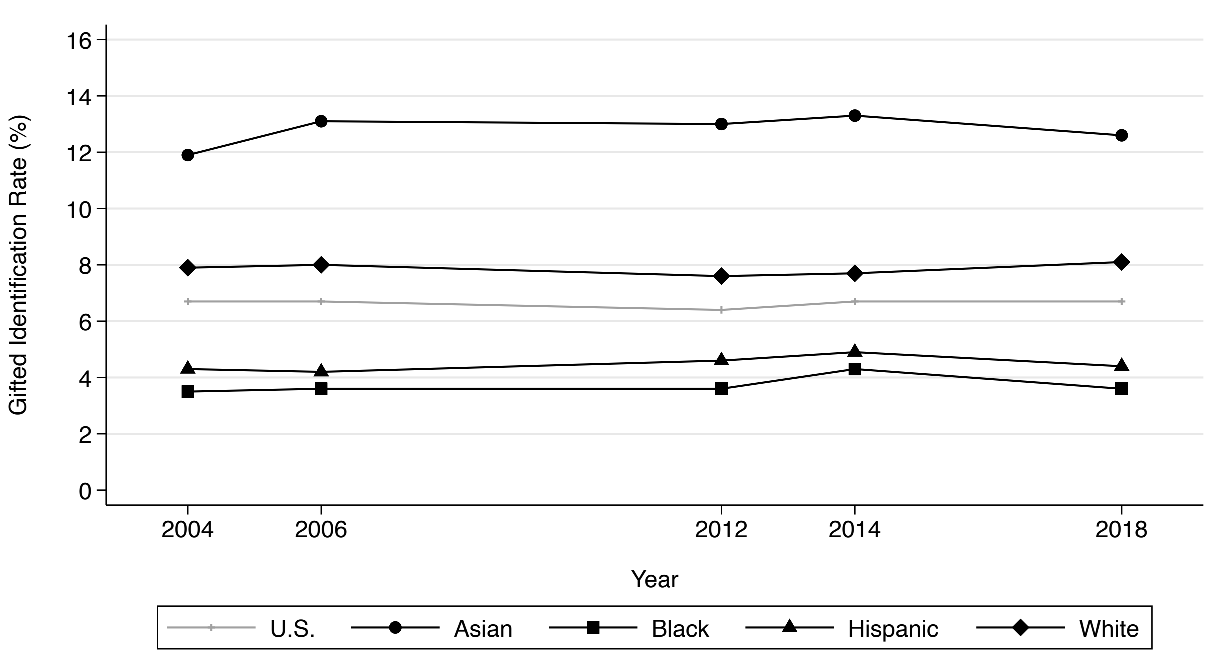
1 Introduction

Gifted and talented (GT) education programs in the United States are designed to deliver curriculum and instruction to high-ability students and serve more than 3 million public school enrollees (Snyder et al., 2020). Proponents of GT programs argue that high-achieving students are insufficiently challenged by existing curricula and deserve resources designed to help them reach their full potential (VanTassel-Baska and Stambaugh, 2007). Qualifying students—who typically access gifted services through some combination of teacher recommendations and demonstrated achievement—can receive a range of interventions, including ability grouping (e.g., Betts and Shkolnik, 2000), course acceleration (e.g., Cohodes, 2020; Hemelt and Lenard, 2020; McCoach et al., 2014), and pullout programming (e.g., Callahan et al., 2015).

A wide range of GT programs appear to confer a number of advantages to participants. Research comparing students who just qualify for GT programs compared to those who do not shows that across a range of subgroups, settings, and outcomes, certain GT education services improve within-district student retention (Davis et al., 2010), grades (Booij et al., 2016, 2017), access to advanced coursework (Backes et al., 2021; Bhatt, 2009), test scores (Cleveland, 2021; Shi, 2020), and college enrollment (Cohodes, 2020; Shi, 2020). These patterns end to hold across a range of subgroups and settings. Additionally, research on tracking initiatives (Card et al., 2016; Duflo et al., 2011), integrated curricula (Callahan et al., 2015; McCoach et al., 2014; Reis et al., 2011), and comprehensive curricula at the secondary level (Lavy and Goldstein, 2022) shows generally positive effects of GT programs. On the other hand, select lottery-based studies of GT participation (e.g., Bui et al., 2014) and enrollment in elite exam schools (Abdulkadiroğlu et al., 2014; Dobbie and Fryer, 2014) do not find meaningful achievement effects or heterogeneity across groups. Moreover, nationwide Evidence from the Early Childhood Longitudinal Study suggests that GT programs neither harmed non-gifted students nor enhanced achievement among gifted students (Adelson et al., 2012) while statewide analyses find that median district effects are not distinguishable from zero (Backes et al., 2021). Thus, findings appear mixed across research methods, forms of GT intervention, and settings.

Although some GT programs benefits appear widespread in the aggregate, they are often not allocated equitably. Gifted education specialists have argued that access to gifted education for traditionally underrepresented students is limited by antiquated identification policies and discriminatory school structures (e.g., Ford, 2003; Morris, 2001; Peters, 2022; Robinson, 2003). Large and persistent disparities in gifted identification exist across demographic groups, notably by race (Imberman, 2021; Shores et al., 2020; Yaluma and Tyner, 2018) and socioeconomic status (Grissom et al., 2019; Imberman, 2021; Yaluma and Tyner, 2018). Federal data show Black and Hispanic students are chronically underrepresented in gifted education programs while white and Asian students are correspondingly overrepresented. Figure 1 highlights this pattern for roughly the past two decades—one that is consistent across the majority of U.S. states (Yoon and Gentry, 2009). One hypothesized reason for these patterns is that referral practices tend to convey patronage based on shared identities, such as race or sex (Elhoweris et al., 2005; Donovan and Cross, 2002; Figlio, 2005; Grissom and Redding, 2016). The shift from such *ad hoc* referrals to universal screeners—where all students test for identification at a common time—can reduce race- and class-based disparities in GT identification (Card and Giuliano, 2015). A second possible reason for persistent gaps is that intervening on behalf of underrepresented students occurs too late in the identification process. Moreover, universal screeners tend to mirror chronic and persistent achievement gaps commonly observed in both low- and high-stakes accountability tests (Reardon, 2011; Reardon et al., 2019). Without cultivating skills and competencies designed to narrow achievement deficits on assessments commonly used for gifted identification, gaps will likely persist.

The purpose of this study is to evaluate an intervention designed to address disparities in GT identification. Nurturing for a Bright Tomorrow (NBT) is an early intervention (grades K-2) with a multi-pronged curriculum centered on developing students' critical thinking skills, intelligent behaviors, and task completion abilities. It differs from most curricular approaches because it targets *all* students—not just potentially gifted qualifiers—early, beginning in kindergarten. The curricular theory of action is that by introducing gifted skills and competencies to the youngest school children across a wide distribution of demographic characteristics and prior achievement, they will be better prepared for the process of formal identification by the third grade. Proponents of curricular revision argue that GT programs should comprise new curricular models accessible to students across demographic, achievement, and income distributions and they should be "frontloaded" early on so that students are exposed to such curricula in advance of formal identification (Olszewski-Kubilius and Clarenbach, 2012; Peters, 2022). NBT does both through its school-level implementation in



Notes: This figure displays gifted identification rates for years in which federal data are available. Each marker represents the proportion of students identified from overall public school enrollment for that group. Rates for 2004, 2006, 2012, and 2014 drawn from annual *Digest of Education Statistics* reports. Rate for 2018 drawn from the Civil Rights Data Collection. Rates for Asian students in 2004 and 2006 include counts for Pacific Islanders.

Figure 1: Public School Gifted Identification Rates, United States, Years Reported from 2004-2018

the early elementary grades.

We implement NBT as a cluster randomized controlled trial across a sample of 32 elementary schools with the lowest gifted identification rates in Wake County Public School System (hereafter, "Wake County"), the largest school district in North Carolina and 15th largest in the nation (**?**). To date, experimental evidence of gifted curricula is limited and suggests that classroom-level curricular interventions generate mixed results (e.g., Callahan et al., 2015; Reis et al., 2011). In work most closely resembling our own, McCoach et al. (2014) study a school-level math enrichment program known as "Project M" and report that the highest performing students at the lowest performing schools benefited most from the curriculum. In contrast to rapid-cycle experiments like Project M, which lasted 16 weeks, Wake County implemented NBT for three years. Thus, students enrolled in treated schools had repeated exposure to the curriculum in the years leading up to formal gifted identification testing. In addition, NBT is subject-agnostic and therefore represents an effort to build underlying gifted competencies that can contribute to downstream content mastery.

We measure the causal impacts of NBT on four sets of outcomes. First, we examine its impact on gifted identification, which students test for in grade 3. Students enrolled in NBT schools received the intervention for three years, from kindergarten to grade 2. Second, we examine NBT's impact on a screener, the Naglieri Nonverbal Abilities Test (NNAT), that the district piloted in order to better understand alternative gifted pathways. Our experiment offered an ideal setting in which to pilot the NNAT with a subsample of students and to test the impact of NBT on this new (to the district) measure. Third, we measure the impact of NBT on math and reading achievement in grades K-2 using the Number Knowledge Test (NKT) and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), which assesses the impacts of NBT on early content knowledge. Finally, we estimate the extent to which NBT fosters engagement by measuring its impacts on excused and unexcused absences.

The impacts of NBT are decidedly mixed. The intervention failed to boost gifted identification rates. In fact, enrollment in an NBT school contributed to slightly lower rates of gifted identification for some groups in some cohorts. This appears to be driven by declines in performance g on the Iowa Assessments (not the CogAT), where the likelihood of meeting or exceeding the 95th percentile declined by 4-9 percentage points. NBT did not boost broad early math or reading achievement, with the exception of moderate gains of roughly 0.09 standard deviations in reading for first graders in the second cohort. Enrollment in an NBT school did substantially increase the likelihood of reaching the gifted threshold on the NNAT, which corresponded to broad achievement improvements in nonverbal ability. As we note above, the district piloted the NNAT and thus did not use it as a formal measure to assess students for GT programs. Still, the results are illuminating. The full sample, along with key subgroups, were 1-2 percentage points more likely to reach the NNAT's published gifted threshold and outperformed control group counterparts by 0.12 standard deviations on the test's nonverbal ability score. These results are consistent across grades and cohorts and are particularly large for Hispanic students and female students.

Our study of NBT contributes to our knowledge of GT interventions, programs, and policies in a few key ways. First, most gifted education settings are designed to introduce a curriculum that meets the needs of *already* high-achieving students (Bhatt, 2011). In contrast, our work introduces a gifted enrichment curriculum to schools with a history of low gifted identification rates. Second, we introduce our curricular experiment early in the education pipeline during grades K-2—a period of development that precedes most formal testing regimes. This approach gives us the opportunity to observe whether gifted skills and competencies can emerge in advance of formal testing in grade 3—an approach referred to as "frontloading" (Peters, 2022). Third, our results potentially generalize to whole school settings—not merely classrooms or clusters of students who may be ability-grouped, which is the case in traditional GT programming. Finally, we implemented NBT for up to three years, which is, to our knowledge, longer than any existing cluster randomized trial designed to boost gifted representation among under-identified groups. Our results paint a mixed picture about the ability of a comprehensive, while-school gifted curriculum to boost outcomes. While Nurturing for a Bright Tomorrow failed to increase early academic achievement the likelihood that students would be identified for GT programming, it did contribute to moderate gains in nonverbal abilities, which include pattern completion, reasoning by analogy, serial reasoning, and spatial visualization—especially among Hispanic and female students.

Below, we describe NBT's curricular components and teacher professional development procedures in Section 2, followed by a description of our setting and data in Section 3 and our analytic plan in Section 4. We then discuss the results in Section 5 and potential mechanisms in Section 6. Finally, we provide a discussion in Section 7 and a conclude in Section 8.

2 Nurturing for a Bright Tomorrow (NBT)

Nurturing for a Bright Tomorrow (NBT) represents an update to what was previously called Project Bright IDEA (Interest Development Early Abilities) and which was motivated by large and persistent gifted education gaps in North Carolina (Darity et al., 2001). NBT 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. In the current iteration, staff from the Academically and Intellectually Gifted (AIG) program in Wake County and AAGS coordinated NBT's implementation. The motivation behind the earlier version—Project Bright IDEA—and the current NBT intervention remains the same: to develop the interests and abilities of traditionally underserved groups in early elementary grades so that by grade 3 they are equipped to qualify for GT programming. The primary difference between the two projects is that while select schools across North Carolina volunteered to implement

Project Bright IDEA, NBT was randomly assigned at the school level within a single, large district. While anecdotal and descriptive evidence suggests that the two previous iterations of Project Bright IDEA—piloted from 2000-2003 and scaled-up from 2004-2009—were positively associated with higher levels of achievement and gifted identification, neither project phase supported causal interpretations of these results (Gayle, 2005; Tzur and Watson, 2010; Watson et al., 2010).

NBT draws from three pedagogical and instructional approaches aimed at developing gifted skills and competencies across the wider student population. The selection of these three components leveraged prior work conducted by NCDPI and AAGS across North Carolina. It also reflected a consensus among various stakeholders that NBT offered enrichment opportunities that far exceeded those provided through off-the-shelf resources (see Section 3) that represented the business-as-usual condition for would-be gifted learners in the early elementary grades.

The first NBT component is aimed at developing students' critical thinking skills. The leading gifted advocacy group, the National Association of Gifted Children (NAGC), argues that to support development of gifted children's strengths in early grades, "we must give them the opportunity to engage in problem solving and employ critical thinking" (O'Brien, 2018). A review of more than 100 studies generally supports this view, as interventions that employ critical thinking strategies show average effects of roughly one-third of a standard deviation on student achievement (Abrami et al., 2008). NBT teachers use the *Building Thinking Skills* book series (Parks and Black, 2012), which includes a wide range of lessons and activities designed to develop these skills (observing, describing, finding similarities and differences, sequencing, and classifying). According to its publisher, the resource 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. NBT Teachers were instructed to integrate the Building Thinking Skills framework into their daily lessons and devote 20 minutes four times weekly to formal coverage of each "skill." A key aspect of this component was to encourage both teachers and students to ask and answer questions in complete sentences.

The second component of NBT draws from philosopher John Dewey's theories of individual disposition, which describe the organization of habits that help individuals—whether teachers (Altan et al., 2019; Dottin, 2009) or students (Kallick and Zmuda, 2017)—to form intelligent behaviors (Dewey, 1923). In *Habits of Mind*, Costa and Kallick (2009) identify 16 such habits that enable children to respond to uncertainty, solve problems creatively, and strengthen social interactions.[1] Former AIG teachers and central office implementation staff developed model units for NBT teachers that emphasized the 16 habits of mind as part of regular classroom instruction in treated schools.

The third and final component of NBT centers on the delivery of differentiated instruction through student completion of increasingly complex tasks. This approach draws from theories of learning styles (e.g., Jung, 1923; Myers, 1962) and the belief that children can develop intellectually when exposed to tasks slightly more difficult than their current abilities might suggest (e.g., Moneta and Csikszentmihalyi, 1996; Vygotsky, 1962). To this end, treated schools implemented the Task Rotations framework (Silver et al., 2007, 2011), which cycles students through discrete tasks aligned with four different learning styles (Mastery Learners, Interpersonal Learners, Understanding Learners, and Self-Expressive Learners). The authors define task rotations as a "framework for differentiating assessment tasks and learning activities so that all students have the opportunity to work in their preferred [learning] styles and to develop their weaker ones." This component includes a wide range of tasks as well as formative and summative end-of-unit assessments to monitor progress.

Taken together, the integrated NBT curriculum combines three theoretically-motivated pedagogical components designed to cultivate critical thinking, intelligent behaviors, and complex task completion. NBT teachers are responsible for learning the three components and implementing them in their classrooms. Below we discuss key aspects of the NBT teachers' professional development.

2.1 Teacher Training and Implementation

While NBT itself is designed to cultivate "gifted" behaviors and competencies, instructional delivery is in the hands of teachers. Prior work suggests that teachers may harbor implicit biases when cultivating gifted potential (e.g., Figlio, 2005; Pearman and McGee, 2022). To help overcome these potential biases, NBT incorporates trainings, classroom observations, post-observation con-

[1]The 16 *Habits of Mind* are applying past knowledge to novel situations, creating, imagining, and innovating, finding humor, gathering data through all senses, listening with understanding and empathy, managing impulsivity, metacognition (thinking about thinking), persisting, questioning and problem posing, remaining open to continuous learning, responding with wonderment and awe, striving for accuracy and precision, taking responsible risks, thinking and communicating with clarity and precision, thinking flexibly, and thinking interdependently.

ferences, and surveys. While most of the implementation team's teacher-level interventions were formative and designed to boost NBT fidelity, we collected survey data at the beginning and end of each school year in order to measure how biases might manifest.

Wake County implemented NBT across three cohorts. The first cohort received NBT by grade-level in a staggered rollout: (1) kindergarten received NBT in 2014-15, (2) 1st grade was added in 2015-16, and (3) 2nd grade was added in 2016-17 (Table 1). Thus, the first cohort consisted of teachers who continued to teach at the same grade level and students who, with the exception of transfers, travelled as a cohort year-by-year. The implementation team (which consisted of staff from the district, Duke University, and guest trainers) delivered the first professional development (PD) and training sessions over a two-day period to kindergarten teachers during the summer of 2014 and to subsequent grade-level teachers during the summers thereafter. Each teaching cohort received refresher PD in subsequent years along with new entrants. For example, the first cohort of kindergarten teachers returned in summer 2015 for a refresher PD while the first cohort of 1st grade teachers and any incoming kindergarten teachers received their initial training. Introductory training and PD sessions consisted primarily of NBT curricular component overviews, model lesson demonstrations, and reflections/discussions among instructors and trainees. Throughout the fall and spring of each academic year, implementation team members conducted multiple day-long sessions to reinforce curricular components. Prior to each school year during summer trainings and the last training session of each academic year, teachers completed a survey that captured individual-level dispositions toward gifted learners. Responses to these surveys helped the implementation team tailor classroom visits and outreach to teachers.

Table 1: NBT implementation across grades, years, and cohorts

	2014-15	2015-16	2016-17	2017-18	2018-19
Grade K	Cohort 1	Cohort 2	Cohort 3		
Grade 1		Cohort 1	Cohort 2	Cohort 3	
Grade 2			Cohort 1	Cohort 2	Cohort 3
Grade 3				Cohort 1	Cohort 2
Grade 4					Cohort 1

Notes: This table shows how NBT was implemented to treatment schools over the intervention period. The intervention period was 2014-15 through 2016-17. Cohorts remain identified, but grayed-out, after this period because Cohorts 2 and 3 tested for gifted identification following the implementation period. For example, Cohort 2 in 2017-18 took the NNAT test as 2nd graders, while Cohorts 1 and 2 tested for gifted identification as 3rd and/or 4th graders.

3 Setting, Data, and Descriptive Statistics

3.1 Setting

Wake County was motivated to launch NBT in the midst of a state-mandated gifted education audit and was required to respond in its cyclical plan about potential interventions.[2] Prior to NBT, the business-as-usual (BAU) condition throughout the district's elementary schools consisted of two resources: (1) Using Science, Talents, and Abilities to Recognize Students—Promoting Learning for Under-Represented Students (USTARS∼PLUS) (Coleman and Shah-Coltrane, 2011) and (2) Primary Education Thinking Skills (PETS) (Nichols et al., 2012; Thomson, 2009). USTARS∼PLUS includes five integrated components designed to help teachers evolve from viewing students as "at risk" to "at potential." The core feature of the program is the Teacher's Observation of Potential in Students (TOPS) instrument, which helps teachers track students' gifted potential. Descriptive evidence suggests that teachers who used USTARS∼PLUS and TOPS were better able to identify high academic ability in traditionally under-identified student groups. PETS is a differentiated learning approach inspired by the school-wide enrichment model (SEM) (Renzulli and Reis, 1985), which teachers are expected to implement at least three times per quarter. Teachers had some discretion about specifically when and how they used these programs and district staff did not systematically monitor implementation fidelity. In contrast, the NBT curriculum was designed to achieve school-level consistency of daily gifted practices in the early elementary grades through the intersection of its three components. Longstanding gifted identification gaps that persisted through the BAU approach motivated the switch to NBT in the experimental sample.

3.2 Data

To construct our analytic sample prior to randomization, we recruited schools that had gifted identification rates below the district's median rate, which yielded 53 elementary schools. To maximize the potential impact of an intervention that would raise rates for all students where

[2]Gifted education in Wake County is currently governed by a statute the North Carolina General Assembly approved in 1996. The statute required local education agencies (LEA) across the state to submit three-year AIG plans beginning in the 1998-99 school year, which means the district was in the midst of its seventh plan during NBT. More recent plans articulated how the district will implement 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."

overall identification was low, we recruited 32 volunteer schools and pairwise matched schools on values of prior school-level gifted identification. We then randomly assigned NBT within pairs (Imbens and Wooldridge, 2009; Imbens and Rubin, 2015). A pairwise randomized experiment represents an extreme form of a stratified randomized experiment, where each of 16 strata contain only one treated unit and one control unit.

Our analytic sample of 32 schools includes roughly 3,500 students in any given cohort-year. We construct demographic indicators using the district's administrative records, which include student-level flags for sex, race/ethnicity, special status (e.g., English Language Learners, students with disabilities) and whether they speak a foreign language at home. Since teachers were charged with implementing NBT, we also use common teacher characteristics drawn from the district's human resources database. The full sample included roughly 200 teachers, which corresponds to an average class size if 17.5.

Table 2 displays descriptive statistics for students and teachers, respectively, prior to randomization. Thus, this sample consists of kindergarten students from the first cohort (among three cohorts) in the fall of the first year of NBT implementation. Columns (1) and (2) display control and treatment means, respectively. The treated sample has slightly more Hispanic students, but this difference between treatment and control samples is not statistically distinguishable from zero. For teachers, those in the experimental sample are more likely to be male and to hold National Board Certification. However, we control for these and other variables in our analysis to improve the precision of our estimates.

3.3 Predictors of Gifted Identification

As a preliminary exercise to ascertain the extent to which such variables may predict the likelihood of students attaining gifted status in the third grade, we fit a series of basic OLS models that account for student characteristics and the characteristics of their teachers. Our model takes the following basic form:

$$GT_i = \alpha + \beta X_i + \epsilon_i, \tag{1}$$

where GT_i is an indicator variable for whether student i qualifies for GT programming in math,

Table 2: Student- and Teacher-Level Balance Statistics

Variable	Control (1)	Treatment (2)	Difference (3)
Student characteristics			
Male	0.510	0.522	0.012
	(0.500)	(0.500)	(0.018)
Asian	0.040	0.038	-0.001
	(0.195)	(0.192)	(0.015)
Black	0.326	0.332	0.006
	(0.469)	(0.471)	(0.062)
Hispanic	0.215	0.277	0.061
	(0.411)	(0.448)	(0.037)
White	0.389	0.324	-0.065
	(0.488)	(0.468)	(0.083)
SWD	0.074	0.080	0.005
	(0.263)	(0.271)	(0.012)
LEP	0.126	0.144	0.018
	(0.331)	(0.351)	(0.032)
Foreign Language at Home	0.169	0.202	0.033
	(0.375)	(0.402)	(0.035)
Observations	1,745	1,831	
Teacher characteristics			
Male	0.010	0.058	0.048**
	(0.100)	(0.235)	(0.022)
Years of Experience	10.190	9.922	-0.268
	(8.452)	(7.806)	(1.064)
National Board Certified	0.140	0.058	-0.082*
	(0.349)	(0.235)	(0.048)
Novice Teacher	0.250	0.204	-0.046
	(0.435)	(0.405)	(0.057)
Same-Race, Non-white	0.170	0.097	-0.073
	(0.378)	(0.298)	(0.055)
Observations	100	103	

Notes: This table displays pre-treatment summary statistics for the first NBT cohort of kindergarten students and kindergarten teachers that entered the sample in fall 2014. Columns (1) and (2) display means and standard deviations for the control and treatment group, respectively, and Column (3) displays differences in means. Each row represents a single regression and standard errors are clustered at the school level, which is the unit of assignment.

reading, or both subjects, β is the coefficient on a vector of student-level characteristics and the characteristics of their teachers, X_i, and ϵ_i is the stochastic error term, which is clustered at the school level. We fit separate models for math, reading, and both subjects—each with and without the inclusion of prior achievement.

Tables 3 and 4 display results for the entire district and a sample of students, respectively, who were third graders at the time of randomization—i.e., a pre-treatment sample unaffected by NBT. The first row of each table displays the sample means of gifted students. First, Table 3 shows that correlates of gifted identification in the district tend to mirror national trends discussed in Section 1. Males are overrepresented in math and underrepresented in reading. Black and Hispanic students are dramatically underidentified in math, reading, and both subjects compared to their white counterparts. Note that prior achievement strongly and consistently predicts gifted identification across models, but even when controlling for this variable (even-numbered columns), the white-Black gifted identification gap persists at 1-2 percentage points. Students with disabilities and English learners are actually more likely to qualify for gifted programming after controlling for prior achievement.

Table 4 shows that as intended, the baseline sample of NBT schools had gifted identification rates that were considerably lower than the district mean. This was especially true among students who qualified for gifted programming in *both* math and reading (Columns 5-6), in which the rate among NBT schools was less than half that of the district: 3% compared to 6.6%. For this sample of schools, the white-Black and white-Hispanic identification gaps are even larger than the district gaps after controlling for prior achievement, at 2-3 percentage points. For the entire district as well as the NBT school sample, characteristics of students' teachers do not correlate with gifted identification. To our surprise, this was true for students of color with teachers of the same race or ethnicity—a relationship that has strong empirical support in a number of studies that explore achievement broadly.[3]

The large documented disparities in identification by race and ethnicity in our wider setting and in the experimental sample further motivated the transition from the district's BAU condition to NBT and its focus on overall impacts by racial/ethnic subgroup as well as sex across subjects.

[3]Blazer (2021), for example, reports large effects when same-race teachers are randomly assigned and includes a comprehensive review of the recent literature.

Table 3: Descriptive Predictors of Gifted Identification in Math and Reading in Grade 3, District Sample

	Math		Reading		Math & Reading	
	(1)	(2)	(3)	(4)	(5)	(6)
Gifted Mean, District	0.039		0.045		0.066	
Student characteristics						
Male	0.032***	0.036***	-0.019***	-0.016***	0.020***	0.030***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Asian	0.103***	0.096***	0.012	-0.015	0.094***	0.050*
	(0.014)	(0.015)	(0.012)	(0.012)	(0.023)	(0.022)
Black	-0.040***	-0.019***	-0.063***	-0.022***	-0.086***	-0.011*
	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.005)
Hispanic	-0.022***	-0.008	-0.038***	-0.015**	-0.057***	-0.013
	(0.007)	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)
Other Race/Ethnicity	-0.012	-0.006	-0.026**	-0.014	-0.010	0.012
	(0.008)	(0.009)	(0.009)	(0.009)	(0.014)	(0.013)
SWD	-0.031***	-0.002	-0.036***	0.020***	-0.050***	0.051***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)
LEP	-0.040***	-0.019*	-0.029***	0.015*	-0.067***	0.010
	(0.009)	(0.008)	(0.006)	(0.006)	(0.012)	(0.010)
Foreign Language at Home	0.012	0.012	-0.009	-0.009	0.013	0.013
	(0.007)	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)
Prior Achievement		0.028***		0.055***		0.100***
		(0.002)		(0.003)		(0.007)
Teacher characteristics						
Male	0.002	0.001	0.010	0.011	-0.009	-0.009
	(0.007)	(0.007)	(0.010)	(0.009)	(0.009)	(0.009)
Years of Experience	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
National Board Certified	0.004	0.004	0.005	0.004	-0.003	-0.010
	(0.006)	(0.006)	(0.008)	(0.008)	(0.007)	(0.007)
Novice Teacher	0.002	0.000	0.001	0.003	-0.002	0.002
	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)
Same-Race, Non-white	-0.003	-0.003	-0.000	0.001	-0.002	0.001
	(0.005)	(0.006)	(0.005)	(0.006)	(0.007)	(0.008)
Observations	12,304	11,615	12,304	11,615	12,304	11,615
Schools	108	108	108	108	108	108

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays predictors of gifted identification for the full sample of district third graders in the year NBT was launched. Thus, it describes a pre-treatment population of qualifiers for illustrative purposes. The first row displays mean gifted identification rates for reference. The coefficients that follow are derived from regressing dichotomous gifted identification in math (Columns (1)-(2)), reading (Columns (3)-(4)), and both subjects (Columns (5)-(6)) on a range of student- and teacher-level characteristics. Even-numbered columns control for prior achievement and have a slightly smaller sample size due to students with missing scores. Standard errors are in parentheses and are clustered at the school level.

Table 4: Descriptive Predictors of Gifted Identification in Math and Reading in Grade 3, Experimental Sample

	Math		Reading		Math & Reading	
	(1)	(2)	(3)	(4)	(5)	(6)
Gifted Mean, NBT School Sample	0.027		0.030		0.030	
Student characteristics						
Male	0.026***	0.032***	-0.014*	-0.008	0.008	0.015
	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)	(0.008)
Asian	0.055	0.054	-0.010	-0.029	0.056	0.024
	(0.028)	(0.029)	(0.016)	(0.017)	(0.028)	(0.017)
Black	-0.039***	-0.023*	-0.056***	-0.025**	-0.061***	-0.024***
	(0.009)	(0.009)	(0.008)	(0.008)	(0.011)	(0.007)
Hispanic	-0.029*	-0.017	-0.036**	-0.019	-0.053***	-0.031***
	(0.012)	(0.011)	(0.011)	(0.011)	(0.009)	(0.006)
Other Race/Ethnicity	-0.001	0.006	-0.025	-0.014	-0.056***	-0.043**
	(0.017)	(0.017)	(0.014)	(0.014)	(0.013)	(0.014)
SWD	-0.026***	-0.001	-0.024***	0.019***	-0.021**	0.030***
	(0.005)	(0.004)	(0.005)	(0.004)	(0.006)	(0.007)
LEP	-0.041**	-0.024	-0.020*	0.013	-0.015	0.026**
	(0.014)	(0.013)	(0.009)	(0.009)	(0.008)	(0.009)
Foreign Language at Home	0.022	0.023	-0.007	-0.004	-0.005	-0.004
	(0.014)	(0.015)	(0.012)	(0.012)	(0.010)	(0.011)
Prior Achievement		0.026***		0.048***		0.055***
		(0.004)		(0.006)		(0.008)
Teacher characteristics						
Male	-0.003	-0.009	0.017	0.015	-0.008	-0.012
	(0.006)	(0.005)	(0.011)	(0.010)	(0.012)	(0.011)
Years of Experience	0.001	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
National Board Certified	-0.005	-0.007	0.007	0.004	0.014	-0.007
	(0.017)	(0.017)	(0.017)	(0.017)	(0.011)	(0.009)
Novice Teacher	-0.008	-0.008	-0.004	-0.004	0.011	0.010
	(0.007)	(0.007)	(0.010)	(0.011)	(0.008)	(0.008)
Same-Race, Non-white	-0.001	-0.002	-0.004	-0.001	-0.001	0.002
	(0.005)	(0.006)	(0.005)	(0.007)	(0.006)	(0.007)
Observations	3,513	3,279	3,513	3,279	3,513	3,279
Schools	32	32	32	32	32	32

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays predictors of gifted identification for the full sample of third graders from NBT schools in the year NBT was launched. Thus, it describes a pre-treatment population of qualifiers for illustrative purposes. The first row displays mean gifted identification rates for reference. The coefficients that follow are derived from regressing dichotomous gifted identification in math (Columns (1)-(2)), reading (Columns (3)-(4)), and both subjects (Columns (5)-(6)) on a range of student- and teacher-level characteristics. Even-numbered columns control for prior achievement and have a slightly smaller sample size due to students with missing scores. Standard errors are in parentheses and are clustered at the school level.

3.4 Outcome Variables

Although Wake County offers a number of pathways to GT programming, the vast majority of students who qualify as gifted do so by meeting or exceeding the 95th percentile on two gifted identification screeners: the Cognitive Abilities Test (CogAt) and the Iowa Assessments (Iowa) during grade 3, when formal testing for gifted identification occurs (See Appendix Figure A1 for a summary of gateways). Thus, we use the district's identification flag based on these assessments to determine whether NBT increases the likelihood of formal identification. To determine whether there is heterogeneity by test, we also present CogAT and Iowa Assessment impacts separately. Finally, for the purposes of this intervention, the district piloted the Naglieri Nonverbal Abilities Test (NNAT) to explore the extent to which it may help boost identification among underrepresented subgroups. The NNAT measures skills and behaviors distinct from what the CogAT and Iowa Assessments capture. Specifically, the NNAT assesses student competencies across four nonverbal domains: pattern completion, reasoning by analogy, serial reasoning, and spatial visualization. NNAT's developers argue that this assessment is better suited than the CogAT or Iowa Assessments for reducing identification disparities (Naglieri and Ford, 2003). In recent years, select large districts have used the NNAT to expand identification procedures (e.g., Bui et al., 2014; Card and Giuliano, 2015, 2016). In this study it serves as an intermediate outcome rather than as a measure of formal identification because it was administered in grades 1 and 2.

In addition to our primary outcome, gifted identification and its components, we measure the impact of NBT on both cognitive and noncognitive outcomes. To estimate cognitive impacts, we rely on data from separate math and reading assessments, which the district administers to all students. To measure early math ability, we use the Number Knowledge Test (NKT) (Okamoto and Case, 1996), which is a screening assessment for early elementary mathematics progress and is designed to measure conceptual knowledge of whole numbers. The raw test score ranges from 1 to 30 and represents grade-level equivalents for pre-school through fifth grade.[4] To measure early reading ability, we use the Dynamic Indicators of Basic Early Literacy (DIBELS). [5] The DIBELS

[4]The NKT has been found to have strong predictive validity for the nationally-normed Stanford Achievement Test (9th edition) and select subtests, with correlations ranging from 0.64 to 0.73 (Baker et al., 2002; Gersten et al., 2005).

[5]According to University of Oregon's Center on Teaching & Learning, the developer, DIBELS "are a set of procedures and measures for assessing the acquisition of early literacy skills from kindergarten through sixth grade... [and] designed to be short (one minute) fluency measures used to regularly monitor the development of early literacy and early reading skills." See https://dibels.uoregon.edu.

composite score ranges from 0 to 564 with benchmark goals for each grade level. For both the NKT and DIBELS, we standardize raw scores to have a mean of zero and unit standard deviation. We summarize our congnitive outcomes in Table 5.

Table 5: Outcomes

Assessment	Grade(s)	Description
Gifted Identification (0/1)	3	The vast majority of students in the district qualify for gifted programming through Gateway #1 (see Figure A1), which requires students to score at or above the 95th percentile on the CogAT and Iowa Assessments.
Cognitive Abilities Test (CogAT) and Iowa Assessments (0/1)	3	The CogAT is the first universal screener students take in order to qualify for gifted programming. If they score at or above the 95th percentile, they are eligible to take the Iowa Assessments. Scoring at or above the 95th percentile in math on both tests, reading in both tests, or math and reading on both tests qualifies students for gifted programming in those respective subject areas.
Naglieri Nonverbal Abilities Test (NNAT) (0/1 and continuous)	1-2	Measures nonverbal reasoning through the identification of shapes, designs, and patterns that are geometrically and/or logically related; represents a popular alternative to the standard gifted identification tests but has no stakes attached in our context and was only administered as a pilot to the analytic sample (i.e., not districtwide). Uses 97.5th percentile as gifted threshold and includes raw scores operationalized here with mean 0 and SD = 1.
Number Knowledge Test (NKT) (continuous)	K-1	Measures counting, number sense, sequences, digits, and basic arithmetic.
Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (continuous)	K-2	Measures letter naming, phonemic awareness, and fluency. Assessments vary by grade level; we use the vendor's standardized composite measure.

Notes: The table above includes a list of outcomes. For more detail on these assessments and relevant citations, see Section 3.4.

We examine the effects of NBT on non-cognitive skill formation and student engagement by using data on absenteeism. This outcome captures features of human capital accumulation that influence students' longer-term performance in ways standardized tests (Jackson, 2018) or participation in gifted programs alone may not capture. We examine excused, unexcused, and total absences, and operationalize this measure in a few ways in order to contextualize absences in ways typically neglected in educational impact evaluations (Gottfried, 2009) or the teacher value-added (Liu and Loeb, 2019) literature.

4 Analytic Strategy

Since NBT was randomly assigned at the school level, we estimate its impacts using a straightforward, intent-to-treat (ITT) OLS specification:

$$Y_{its} + \alpha + \beta NBT_j + \gamma \boldsymbol{X}_i + \pi_b + \varepsilon_{its}. \tag{2}$$

The main effect of the offer to implement NBT on any outcome Y is given by β and represents an ITT effect. The row vector X includes pre-treatment covariate values for student demographic characteristics (e.g., race/ethnicity, sex, disability status, English learner status), pre-treatment assessment measures, characteristics of students' teachers, and randomization blocks (π_b) that result from matching schools on previous measures of gifted identification prior to implementation. Standard errors are clustered at the school level, which is the unit of randomization.

Equation 2 represents the simplest form of the model we use to estimate treatment effects on student outcomes for a single time period. For example, this model would estimate treatment effects for students in a single cohort in a single year, such as students in Cohort 1 who are assessed for gifted identification as third graders in spring 2017. For models that include multiple cohorts over multiple years, we pool observations and specify the above model with the inclusion of year, grade, and cohort fixed effects.

5 Results

5.1 Gifted Identification

The results in Table 6 address the primary question of this study: Did NBT boost the likelihood that students would qualify for GT programs in grade 3? The table displays results for gifted identification in math, reading, and both subjects in columns and heterogeneity across rows. Panel A displays the main results. As we note in Section 3.4, students may qualify for gifted programming in either math, reading, or both subjects. Reading across Panel A, we focus on the pooled results in Columns (3), (6), and (9). In the main sample and for empirically relevant subgroups in Panels B-E, Column (3) shows that NBT did not impact the likelihood of identifying as gifted in math. The same is largely true in reading, however Hispanic students experienced an increase of one-half of a percentage point in reading. Finally, NBT appears to have negatively impacted the likelihood of qualifying for gifted programing in both math and reading among Black and female students. These negative impacts are substantively meaningful and are driven by results for Cohort 1, which received the greatest dosage (3 years).

Table 6: Impacts of Nurturing for Bright Tomorrow (NBT) on Qualifying for Gifted Identification. Grade 3

	Gifted in Math			Gifted in Reading			Gifted in Both		
	Grade 3			Grade 3			Grade 3		
	Cohort 1	Cohort 2	Pooled	Cohort 1	Cohort 2	Pooled	Cohort 1	Cohort 2	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Full Sample									
NBT	-0.003	-0.003	-0.003	0.006	-0.007	-0.000	-0.013*	-0.000	-0.007
	(0.005)	(0.006)	(0.003)	(0.004)	(0.004)	(0.003)	(0.006)	(0.008)	(0.005)
Observations	3292	3014	6306	3292	3014	6306	3292	3014	6306
Panel B: Black									
NBT	0.007	0.004	0.006	-0.001	-0.003	-0.001	-0.009**	-0.004	-0.007**
	(0.006)	(0.007)	(0.004)	(0.005)	(0.006)	(0.004)	(0.003)	(0.005)	(0.002)
Observations	1088	975	2063	1088	975	2063	1088	975	2063
Panel C: Hispanic									
NBT	-0.018*	-0.002	-0.010	0.005	0.004	0.005*	-0.009	0.003	-0.003
	(0.007)	(0.007)	(0.005)	(0.004)	(0.003)	(0.002)	(0.008)	(0.005)	(0.004)
Observations	817	776	1593	817	776	1593	817	776	1593
Panel D: Female									
NBT	0.003	-0.002	-0.001	0.011	-0.004	0.003	-0.028**	-0.003	-0.016*
	(0.005)	(0.006)	(0.004)	(0.009)	(0.006)	(0.006)	(0.009)	(0.011)	(0.007)
Observations	1601	1481	3082	1601	1481	3082	1601	1481	3082
Panel E: Male									
NBT	-0.006	-0.005	-0.006	0.002	-0.008	-0.003	-0.000	0.002	0.000
	(0.008)	(0.008)	(0.004)	(0.004)	(0.006)	(0.003)	(0.009)	(0.008)	(0.007)
Observations	1691	1533	3224	1691	1533	3224	1691	1533	3224

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays treatment effects for NBT on the likelihood of qualifying for gifted programming. Columns (1)-(3) display results for math for Cohort 1, Cohort 2, and for both cohorts pooled. Columns (4)-(6) display these results for reading. Columns (7)-(9) display results for students gifted in both subjects. All coefficients are estimated via linear probability models and presented as percentage point changes. In the most common gifted pathway, students scoring at or above the 95th percentile on the CogAt qualify to take the Iowa Assessments. All models include fixed effects for randomization matched pairs and pooled models include cohort and year fixed effects. Standard errors are clustered at the school level, which is the level of randomization.

What potentially drives the broad, subject-level null effects and dual-subject negative effects for Black and female students? Tables 7 and 8 present test heterogeneity for the two exams that underlie formal gifted identification in the district. These two tables are oriented similarly to Table 6. Panel A shows that NBT led to a 1.2 percentage point increase in the likelihood of the pooled sample identifying for gifted programming in math. This represents nearly half of the sample mean seen in Table 4. There were no corresponding treatment effects for any subgroups pooled by cohort, suggesting that students in treated schools were, on average, as likely to reach the CogAt threshold required to qualify to take the Iowa Assessments. Here, major differential impacts emerge. Table 8 shows that in the full sample, students were 4-9 percentage points less likely to reach the Iowa threshold. These declines were comparable across genders in math and even larger for Hispanic students. In reading, a subject areas where female students broadly outperform males, females were 7 percentage points less likely than their control group counterparts to reach the Iowa qualification threshold. One might be concerned about imbalance among Iowa examinees, since qualification by treatment condition may differentially vary. However, Appendix Table A1 shows that there is strong balance between treatment and control groups, suggesting that this Iowa subsample is comparable to the CogAT sample.

Table 7: Impacts of Nurturing for Bright Tomorrow (NBT) on Meeting 95th Percentile on CogAT, Grade 3

	CogAT Math			CogAT Reading			CogAT Nonverbal		
	Grade 3			Grade 3			Grade 3		
	Cohort 1	Cohort 2	Pooled	Cohort 1	Cohort 2	Pooled	Cohort 1	Cohort 2	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Full Sample									
NBT	0.013	0.012	0.012*	0.001	-0.001	-0.000	-0.001	0.008	0.003
	(0.007)	(0.007)	(0.005)	(0.004)	(0.006)	(0.003)	(0.007)	(0.007)	(0.005)
Observations	2838	2528	5366	2985	2653	5638	3023	2703	5726
Panel B: Black									
NBT	0.009	0.007	0.008	0.002	-0.003	0.001	-0.001	-0.002	-0.003
	(0.006)	(0.005)	(0.004)	(0.004)	(0.006)	(0.004)	(0.008)	(0.007)	(0.004)
Observations	893	763	1656	964	821	1785	983	837	1820
Panel C: Hispanic									
NBT	-0.008	0.000	-0.004	0.006*	-0.003	0.001	-0.003	0.013*	0.007
	(0.006)	(0.006)	(0.003)	(0.002)	(0.004)	(0.003)	(0.009)	(0.005)	(0.005)
Observations	710	644	1354	726	671	1397	752	688	1440
Panel D: Female									
NBT	-0.001	0.010	0.004	-0.015	0.008	-0.003	-0.023*	0.019	-0.003
	(0.012)	(0.007)	(0.006)	(0.007)	(0.010)	(0.005)	(0.011)	(0.010)	(0.007)
Observations	1367	1240	2607	1464	1324	2788	1481	1349	2830
Panel D: Male									
NBT	0.026	0.011	0.019*	0.015	-0.013	0.002	0.019*	-0.001	0.008
	(0.014)	(0.010)	(0.009)	(0.008)	(0.008)	(0.004)	(0.007)	(0.009)	(0.005)
Observations	1471	1288	2759	1521	1329	2850	1542	1354	2896

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays treatment effects for NBT on the likelihood of scoring at or above the 95th percentile on the CogAT Assessments. Columns (1)-(3) display results for math for Cohort 1, Cohort 2, and for both cohorts pooled. Columns (4)-(6) display these results for reading. All coefficients are estimated via linear probability models and presented as percentage point changes. In the most common gifted pathway, students scoring at or above the 95th percentile on the CogAt qualify to take the Iowa Assessments. All models include controls for student demographics, special classifications, and teacher characteristics, and fixed effects for randomization matched pairs. Pooled models include cohort and year fixed effects. Standard errors are clustered at the school level, which is the level of randomization.

Table 8: Impacts of Nurturing for Bright Tomorrow (NBT) on Meeting 95th Percentile on Iowa Assessments, Grade 3

	Iowa Math			Iowa Reading		
	Grade 3			Grade 3		
	Cohort 1	Cohort 2	Pooled	Cohort 1	Cohort 2	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full Sample						
NBT	-0.129**	-0.051	-0.089**	-0.023	-0.074**	-0.042*
	(0.036)	(0.034)	(0.030)	(0.025)	(0.026)	(0.020)
Observations	985	876	1861	986	877	1863
Panel B: Black						
NBT	-0.037	-0.094	-0.043	0.018	0.066	0.038
	(0.041)	(0.048)	(0.034)	(0.042)	(0.037)	(0.028)
Observations	210	193	403	210	192	402
Panel C: Hispanic						
NBT	-0.162***	-0.077	-0.118**	-0.015	-0.034	-0.014
	(0.031)	(0.090)	(0.035)	(0.047)	(0.046)	(0.033)
Observations	157	126	284	157	126	284
Panel D: Female						
NBT	-0.115*	-0.063	-0.084*	-0.046	-0.123*	-0.071*
	(0.051)	(0.033)	(0.035)	(0.040)	(0.047)	(0.034)
Observations	455	403	858	455	403	858
Panel E: Male						
NBT	-0.152***	-0.037	-0.095**	-0.015	-0.039	-0.021
	(0.039)	(0.049)	(0.030)	(0.024)	(0.028)	(0.017)
Observations	530	473	1003	531	474	1005

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays treatment effects for NBT on the likelihood of scoring at or above the 95th percentile on the Iowa Assessments. Columns (1)-(3) display results for math for Cohort 1, Cohort 2, and for both cohorts pooled. Columns (4)-(6) display these results for reading. All coefficients are estimated via linear probability models and presented as percentage point changes. Students scoring at or above the 95th percentile on the CogAt qualify to take the Iowa Assessments. All models include controls for student demographics, special classifications, and teacher characteristics, and fixed effects for randomization matched pairs. Pooled models include cohort and year fixed effects. Standard errors are clustered at the school level, which is the level of randomization.

The results in Tables 6-8 suggest that, broadly, schools utilizing the BAU gifted preparation resources (USTARS∼PLUS and PETS) were considerably more successful in qualifying for gifted programming. This result appears hinge on the ability of students in non-NBT schools to outperform their treated counterparts on the gateway Iowa Assessments. Moreover, while not directly testable, the Iowa result suggests that BAU condition may be more impactful due to its relatively targeted approach compared to NBT, which represents a universal intervention (Ceci and Papierno, 2005) that is spread too thinly across the population.

In addition to measuring the impact of NBT on formal gifted identification and the assessments that underlie it, the district piloted the Naglieri Nonverbal Abilities Test (NNAT), which its developers argue is a more appropriate assessment for English language learners, students living in poverty, or those with low levels of academic achievement (Naglieri and Ford, 2003). The theory supporting this claim hinges on the assumption that students from a diverse range of backgrounds can perform visual, "progressive matrix" tasks—e.g., identify the missing piece in a series of shapes—without the need to internalize cultural or linguistic norms. Table 9, Panel A, shows that students enrolled in NBT schools were 1.5 percentage points more likely to reach the NNAT's gifted threshold (Column 4). This effect corresponded to effect size gains of 0.12 standard deviations based on standardized raw scores that provide more granularity than the percentile ranks used for identification. Scanning down Column (4) reveals that identification effects appear driven by effects among Hispanic students and female students. The corresponding test score effect sizes for the full sample and these two groups ranged from 0.11-0.19 *SD*s, which are considered empirically moderate effects in the context of school-based randomized experiments with student-level outcomes (Kraft, 2020; Lipsey et al., 2012).

Table 9: Impacts of Nurturing for Bright Tomorrow (NBT) on Nonverbal NNAT Outcomes

	Identification Likelihood (*pp*)				Scale Score Effect Size (*SDs*)			
	Grade 1 Cohort 1	Grade 2 Cohort 1	Grade 2 Cohort 2	Pooled	Grade 1 Cohort 1	Grade 2 Cohort 1	Grade 2 Cohort 2	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Full Sample								
NBT	0.007	0.018*	0.017***	0.015*	0.064	0.144**	0.127**	0.119***
	(0.007)	(0.007)	(0.004)	(0.005)	(0.036)	(0.041)	(0.045)	(0.033)
Observations	2706	2978	2860	8544	2706	2978	2860	8544
Panel B: Black								
NBT	-0.002	0.002	0.013*	0.003	-0.067	0.057	0.077	0.030
	(0.008)	(0.006)	(0.006)	(0.004)	(0.041)	(0.071)	(0.066)	(0.046)
Observations	873	983	862	2718	873	983	862	2718
Panel C: Hispanic								
NBT	0.006	0.011*	0.011*	0.011*	0.150*	0.160*	0.199**	0.186**
	(0.010)	(0.004)	(0.004)	(0.004)	(0.066)	(0.076)	(0.065)	(0.055)
Observations	687	767	714	2168	687	767	714	2168
Panel D: Female								
NBT	0.016	0.021	0.017*	0.017*	0.050	0.152*	0.194***	0.136**
	(0.013)	(0.011)	(0.006)	(0.008)	(0.064)	(0.055)	(0.051)	(0.043)
Observations	1315	1454	1391	4160	1315	1454	1391	4160
Panel E: Male								
NBT	-0.008	0.017*	0.017**	0.011*	0.037	0.116**	0.051	0.080*
	(0.007)	(0.008)	(0.006)	(0.004)	(0.036)	(0.040)	(0.053)	(0.031)
Observations	1391	1524	1416	4331	1391	1524	1416	4331

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays treatment effects for NBT on two sets of Naglieri Nonverbal Ability Test (NNAT) outcomes. Columns (1)-(4) display percentage point likelihoods of treated students qualifying for the NNAT gifted threshold by scoring at or above the 97.5th percentile. Columns (5)-(8) show corresponding effect sizes based on scale scores for the same assessment. Note that the NNAT was administered as an additional, pilot assessment to the district's traditional battery of gifted identification exams (i.e., CogAt and Iowa) and did not formally qualify students for formal entry into gifted programming. All coefficients are estimated via linear probability models. All models include controls for student demographics, special classifications, and teacher characteristics, and fixed effects for randomization matched pairs. Pooled models include cohort and year fixed effects. Standard errors are clustered at the school level, which is the level of randomization.

5.2 Academic Achievement

To estimate achievement effects, we use the NKT for math and DIBELS for reading, which are administered three times annually. We show the effects of NBT at the conclusion of the intervention using each end-of-year score as the outcome variable and the student's first observable pre-treatment fall score. Table 10, Panel A, shows that NBT's effects on math and reading for the pooled cohorts were opposite-signed but noisily estimated. Scanning down Columns (3) and (7) reveals that these imprecise null effects are consistent across relevant subgroups. Within cohorts and subgroups, no discernible pattern emerges. If anything, Black and female students in the last cohort experienced declines in math while these same groups from later cohorts experienced gains in reading. If anything, we conclude that NBT did not impact traditional cognitive measures in ways that might parallel effects seen on the Iowa Assessments or NNAT.

Table 10: Impacts of Nurturing for Bright Tomorrow (NBT) on Math and Reading Achievement

	NKT Math			DIBELS Reading			
	Kindergarten Cohort 3	Grade 1 Cohort 2	Pooled	Kindergarten Cohort 3	Grade 1 Cohort 2	Grade 2 Cohort 1	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Full Sample							
NBT	-0.098	0.026	-0.034	0.049	0.094*	-0.034	0.037
	(0.067)	(0.085)	(0.060)	(0.048)	(0.040)	(0.050)	(0.035)
Observations	2366	2756	5122	3179	3307	3385	9871
Panel B: Black							
NBT	-0.159*	-0.137	-0.125	0.025	0.117**	-0.107	0.028
	(0.074)	(0.115)	(0.086)	(0.056)	(0.039)	(0.074)	(0.033)
Observations	819	910	1729	994	1064	1134	3192
Panel C: Hispanic							
NBT	-0.074	0.141	0.045	0.098	0.093	0.057	0.089
	(0.054)	(0.096)	(0.054)	(0.060)	(0.093)	(0.068)	(0.062)
Observations	677	702	1379	785	822	857	2464
Panel D: Female							
NBT	-0.129*	-0.015	-0.063	0.108*	0.065	-0.079	0.028
	(0.060)	(0.090)	(0.060)	(0.052)	(0.056)	(0.059)	(0.042)
Observations	1162	1322	2484	1543	1605	1657	4805
Panel D: Male							
NBT	-0.081	0.030	-0.029	-0.011	0.084	-0.026	0.026
	(0.084)	(0.090)	(0.064)	(0.057)	(0.056)	(0.053)	(0.040)
Observations	1204	1434	2638	1636	1702	1728	5066

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays treatment effects of NBT on standardized achievement scores on the Number Knowledge Test (NKT) and Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assessment. NKT results are displayed in Columns (1)-(3) and are read left to right by increasing grade level exposure to treatment. Second graders in Cohort 1 at the conclusion of NBT were not administered the NKT. DIBELS results are displayed in Columns (4)-(7) and are also read in order of increasingly grade level. All models include controls for student demographics, special classifications, and teacher characteristics, and fixed effects for randomization matched pairs. Pooled models include cohort and year fixed effects. Standard errors are clustered at the school level, which is the level of randomization.

5.3 Absenteeism

We hypothesize that students attending NBT schools would experience a decline in absenteeism due to the various approaches to student engagement woven throughout the specialized curriculum. We measure absences in three ways: (1) excused, (2) unexcused, and (3) total. We operationalize absenteeism using the log of absences plus 1, a dichotomous indicator for any absenteeism, and chronic absenteeism. To our surprise, the intervention did not reduce absenteeism, but rather increased the incidence of excused absences across all three cohorts (Appendix Table A2.) Consistent with our hypothesis, but amid mostly null effects, NBT reduced chronic absenteeism by 2 percentage points for Cohort 3. While the increase in excused absenteeism is is puzzling, previous work from Gottfried (2009) suggests a positive association between excused absences and standardized test performance in reading and math. He attributes this result to highly motivated students obtaining legitimate reasons for absenteeism and parents who permit such absences while opposing unexcused absences that signal truancy. Indeed, all nine coefficients on unexcused absences are negative, though nearly all are highly imprecise.

6 Potential Mechanisms

6.1 Implementation

We summarize implementation procedures in Section 2.1. Most procedures were formative and designed to improve implementation where it was weak. Members of the district's AIG team visited schools roughly weekly to observe classrooms and confer with teachers over various components of the implementation checklist. We did not code the results from these checklists because they were not consistent across classrooms, grades, and implementation years. We can, however, draw three qualitative conclusions based on informal teacher feedback. First, teachers were overwhelmed by the implementation of three new curricular resources—each of which merited separate training sessions. While NBT was designed as a synthesized intervention that threads all three components throughout the day, training during the first year did not reflect this approach. Second, training in the subsequent year—which was outsourced to external trainers—received low marks from teachers. This made it difficult to generate sufficient buy-in from second-year participants. Finally, in response to feedback from Cohort 1 and 2 teachers to reduce training—given their existing exposure to the curriculum—the NBT implementation team consolidated modules. Unfortunately, this came at the expense of Cohort 3 kindergarten teachers, who were brand new to the curriculum.

In an effort to imperfectly quantify implementation after NBT's third year, we put three prompts to teachers asking their level of agreement regarding the impact of NBT on their own pedagogy, their students' thinking skills, and PD more generally. Each column of Table 11 presents associations between teacher characteristics and perception of implementation fidelity captured by these three prompts. The percentage of respondents who agreed or strongly agreed with each statement appear in the penultimate row, and ranges from 40-64%. The only teacher-level characteristics associated with perceptions of implementation was whether the respondent possessed National Board certification. Surprisingly, National Board-certified teachers were considerably more likely to agree that NBT impacted their teaching and their students learning, but were considerably less likely to agree that professional development was valuable. Such a result is consistent with our intuition that PD was delivered with mixed quality over the three year intervention—an outcome recognized by these relatively motivated educators who complete the rigorous National Board certification process. While this exercise only presents a fraction of self-reports, this data collection method for measuring implementation fidelity is fairly common and, given that only roughly half of respondents agreed that NBT was beneficial, suggests a potential link between implementation fidelity and outcomes (Hill and Erickson, 2019).

6.2 Teacher Disposition

Previous evidence suggests that one major factor contributing to gifted identification gaps is educator implicit bias (Figlio, 2005; Grissom and Redding, 2016). Since NBT largely failed to achieve the primary goals of boosting rates of gifted identification among traditionally under-identified groups, we hypothesize that such biases may have played a role. To shed light on the possible links between biases and outcomes, we surveyed teachers throughout the intervention period about such views using a 53-item "teacher disposition" survey. This survey was introduced by developers of Project Bright IDEA—NBT's precursor. To simplify the survey, we reduced the number of items using a theory-driven approach (Kraft et al., 2016) that resulted in 32 items (see Appendix A.1) that retained strong internal consistency (Chronbach's alpha = 0.81). In completing the survey, teach-

Table 11: Teacher Likelihood of Agreeing or Strongly Agreeing with Statements Related to the Quality of NBT Implementation

	Nurturing for a Bright Tomorrow . . .		
	Has impacted me as a teacher	Improved my students' thinking skills	Provided effective professional development
	(1)	(2)	(3)
Male	-0.283	0.174	-0.210
	(0.373)	(0.247)	(0.125)
Nonwhite	0.124	-0.036	0.209
	(0.183)	(0.180)	(0.180)
Years of Experience	-0.016	-0.007	-0.003
	(0.010)	(0.014)	(0.008)
National Board Certified	0.534**	0.314*	-0.756***
	(0.141)	(0.120)	(0.121)
Novice	-0.082	-0.010	0.073
	(0.202)	(0.227)	(0.222)
Mean agreement (%)	57.81	64.06	39.68
Observations	63	63	63

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays associations between teacher characteristics and likelihood of agreeing or strongly agreeing with statements related to the quality of NBT implementation. Models are specified as OLS with dichotomous dependent and independent variables, with the exception of the teacher experience variable, which is continuous.

ers were forced to consider their own role, student ability, and parental support. Prompts included "Academic giftedness depends on a teacher's nurturing effort," "Students' unique racial background is an important resource in my planning for instruction," and "A teacher should help parents form realistic expectations about their child's giftedness." As with the implementation prompts, teachers responded using a 5-point Likert scale ranging from "strongly disagree" to "strongly agree." To ascertain the extent to which teacher disposition varied between those in the treatment and control group by the end of NBT, we regressed a dichotomous level of agreement on the treatment indicator and conditioned on teacher characteristics. This resulted in statistically significant difference on one item (item 12, $p < 0.05$) and suggestive differences on two items (items 20 and 32, $p < 0.10$), with results that do not point to a discernible role for implicit biases. To be sure, our respondent sample of 122 results is underpowered. However, a similar number of items are either negatively or positively estimates, suggesting an ambiguous influence of disposition.

7 Discussion

NBT's mixed impacts highlight the complexity associated with reducing disparities in gifted identification. To start, our setting, like many others across the U.S., has large and persistent gifted identification gaps between white students and their Black and Hispanic counterparts. The district's longstanding BAU condition has done very little to narrow those gaps. This fact, along with a districtwide gifted education audit ordered by the state prior to NBT, motivated the curriculum's implementation. NBT introduced a whole-school curriculum in grades K-2 that was designed to boost gifted competencies in schools with chronically low gifted identification rate. However, the implementation and integration of NBTs three components did not boost identification by grade 3, suggesting that the BAU condition, at minimum, did no harm.

Beyond the core curricular components, NBT's approach also included piloting a new (to the district) assessment that developers argued was better suited for identifying traditionally underidentified students and measuring the extent to which teacher biases might exacerbate gaps. To date, we still have limited understanding about whether one such assessment, the NNAT, can identify gifted competencies among such groups (e.g., Lohman, 2005; Naglieri and Ford, 2005; Giessman et al., 2013). Carman et al. (2020) show that neither the NNAT nor the CogAt Verbal battery are better able to identify traditionally underidentified students. Our study suggests that in combination with NBT, it does—and by a substantively meaningful degree representing roughly a third of the average identification rate in our analytic sample. Still, critics of the NNAT as a primary means of identification (e.g., Lohman, 2005; Giessman et al., 2013) argue that educators who focus on nonverbal, spatial, and figural reasoning do so at the expense of verbal and quantitative development—a tradeoff that could, in fact, depress identification, and which might help explain poor performance on the gateway Iowa Assessments. While NBT was assessment agnostic, an unintended consequence of this curricular pivot may have overemphasized new domains at the expense of domains traditionally assessed for gifted identification. Moreover, the universal nature of the intervention as a school-level experiment suggests that a targeted approach may have been more beneficial (Ceci and Papierno, 2005).

Likewise, we have limited evidence demonstrating the extent to which students assigned to same-race teachers are more likely to identify as gifted. Grissom and Redding (2016) find that when Black students are assigned to a Black teacher, gifted identification disparities narrow. In our setting, same-race assignment does not predict variation in identification and the role of teacher biases appear ambiguous. Still, future research should further probe the extent to which students believe their teachers hold implicit biases and whether teachers themselves harbor such biases toward potential gifted learners.

While we were not able to document mechanisms, we do recognize that implementation was uneven. We know from semi-structured interviews with school staff that there existed considerable variation in levels of implementation. This sentiment was partially validated through a review of the teacher disposition survey results, which pointed to mixed impressions of the program's impact and no meaningful reduction of implicit bias in treated schools. As Hill and Erickson (2019) confirm in meta-analytic work, implementation and effects are strongly related, and even moderately implemented interventions can generate positive results. Based on implementation data discussed in Section 6.1, we believe NBT was at least moderately implemented, which can help explain the NNAT result in light of frequent null and occasional mixed impacts.

8 Conclusion

Gifted identification disparities remain a vexing challenge in U.S. public education. We do not know why such gaps exists or persist, but we have some ideas.[6] For instance, subjective nomination practices steeped in racial and socioeconomic biases potentially lead to white, affluent students being overrepresented in gifted programs. In addition, the federal law governing gifted education does not include AG program implementation mandates, which results in decentralized and heterogeneous program designs that may benefit some student groups over others. The intervention we study, Nurturing for a Bright Tomorrow (NBT), aimed to address these suspect causes in a number of ways. First, in addition to the universal screeners implemented district wide, NBT used a separate screener that many argue is more appropriate for identifying gifted competencies in underrepresented groups (e.g., Naglieri and Ford, 2003). Second, NBT was implemented school-wide and was therefore centralized and implemented in all K-2 classrooms. Finally, the professional development associated

[6] Peters (2022), introducing a special issue of *Gifted Child Quarterly* devoted to explanations of disparities in gifted identification, argues that predictors include poverty, lead exposure, adverse childhood experiences, and police violence.

with NBT was designed to identify and address educator biases that may establish barriers to gifted education for underrepresented student groups.

Unfortunately, the district's curricular pivot for early elementary students did not, on average, impact students as hypothesized. With the exception of some benefits for Hispanic students, NBT did not broadly boost identification rates or achievement. However, it did improve outcomes on the nonverbal pilot assessment that, in combination with a whole-school curriculum like NBT, holds promise for narrowing racial/ethnic gifted disparities. Our work is not without limitations. First, our cluster randomized trial was largely underpowered to detect average and heterogeneous effects. Still, since we had an opportunity to study a policy-relevant equity issue through a strong district-university partnership, we proceeded despite the relatively small number of clusters (Angrist and Pischke, 2009, p. 319). Second, training for the implementers—K-2 teachers—was delivered with mixed consistency and rigor. Finally, despite frequent school visits, observations, and consultations that guided formative implementation meetings, we did not sufficiently capture quantitative implementation data that could have helped better contextualize our results. Future work should continue to identify interventions, gateways, and assessment batteries that reduce frictions for underrepresented student groups on the path to gifted and talented educational opportunities.

The mixed results present a puzzle. Nurturing for a Bright Tomorrow was designed to cultivate gifted behaviors that are thought to increase one's chance of qualifying for gifted programming. Specifically, NBT's integrated curriculum combined critical thinking, problem solving behaviors, and complex task completion as a way to narrow disparities. Yet the concentrated development of these skills in treatment classrooms did not lead to higher identification rates underpinned by cognitive test scores. Instead, it led to higher nonverbal abilities. The results suggest that the NBT curricular approach enhanced an ability type not captured by longstanding, traditional measures of cognitive attainment. Although the results verge on disappointing, they also highlight the extent to which traditional gifted identification procedures may perpetuate chronic identification gaps. The mixed-to-negative impacts on status quo measures of identification and achievement demonstrate how difficult it is to move the needle when the identification paradigm is steeped in long-established measures of cognitive attainment. On the other hand, the moderate gains in non-verbal ability captured by the NNAT suggest that less-frequently measured gifted competencies can be captured through alternative screening procedures.

References

Abdulkadiroğlu, A., Angrist, J., and Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, 82(1):137–196.

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., and Zhang, D. (2008). Instructional Interventions Affecting Critical Thinking Skills and Dispositions: A Stage 1 Meta-Analysis. *Review of Educational Research*, 78(4):1102–1134.

Adelson, J. L., McCoach, D. B., and Gavin, M. K. (2012). Examining the Effects of Gifted Programming in Mathematics and Reading Using the ECLS-K. *Gifted Child Quarterly*, 56(1):25–39.

Altan, S., Lane, J. F., and Dottin, E. (2019). Using Habits of Mind, Intelligent Behaviors, and Educational Theories to Create a Conceptual Framework for Developing Effective Teaching Dispositions. *Journal of Teacher Education*, 70(2):169–183.

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Backes, B., Cowan, J., and Goldhaber, D. (2021). What Makes for a" Gifted" Education? Exploring How Participation in Gifted Programs Affects Students' Learning Environments. *CALDER Working Paper No. 256-0821*.

Baker, S., Gersten, R., Flojo, J., Katz, R., Chard, D., and Clarke, B. (2002). Preventing mathematics difficulties in young children: Focus on effective screening of early number sense delays. Technical report, Technical Report.

Betts, J. R. and Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review*, 19(1):1–15.

Bhatt, R. (2011). A Review of Gifted and Talented Education in the United States. *Education Finance and Policy*, 6(4):557–582.

Bhatt, R. R. (2009). The Impacts of Gifted and Talented Education. *SSRN Electronic Journal*.

Blazer, D. (2021). Teachers of Color, Culturally Responsive Teaching, and Student Outcomes: Experimental Evidence from the Random Assignment of Teachers to Classes.

Booij, A., Haan, F., and Plug, E. (2016). Enriching Students pays off: Evidence from an Individualized Gifted and Talented Program in Secondary Education. *IZA Discussion Paper No. 9757*, page 37.

Booij, A., Haan, F., and Plug, E. (2017). Can gifted and talented education raise the academic achievement of all high-achieving students? page 29.

Bui, S. A., Craig, S. G., and Imberman, S. A. (2014). Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Students. *American Economic Journal: Economic Policy*, 6(3):30–62.

Callahan, C. M., Moon, T. R., Oh, S., Azano, A. P., and Hailey, E. P. (2015). What Works in Gifted Education: Documenting the Effects of an Integrated Curricular/Instructional Model for Gifted Students. *American Educational Research Journal*, 52(1):137–167.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *The Quarterly Journal of Economics*, 131(2):633–686.

Card, D. and Giuliano, L. (2015). Can Universal Screening Increase the Representation of Low Income and Minority Students in Gifted Education? Technical Report w21519, National Bureau of Economic Research, Cambridge, MA.

Card, D. and Giuliano, L. (2016). Can Tracking Raise the Test Scores of High-Ability Minority Students? *American Economic Review*, 106(10):2783–2816.

Carman, C. A., Walther, C. A. P., and Bartsch, R. A. (2020). Differences in Using the Cognitive

Abilities Test (CogAT) 7 Nonverbal Battery Versus the Naglieri Nonverbal Ability Test (NNAT) 2 to Identify the Gifted/Talented. *Gifted Child Quarterly*, 64(3):171–191.

Ceci, S. J. and Papierno, P. B. (2005). The Rhetoric and Reality of Gap Closing: When the "Have-Nots" Gain but the "Haves" Gain Even More. *American Psychologist*, 60(2):149–160.

Cleveland, C. (2021). The Impacts of Gifted and Talented Education Program Participation: Evidence from New York City. *Working paper*.

Cohodes, S. R. (2020). The Long-Run Impacts of Specialized Programming for High-Achieving Students. *American Economic Journal: Economic Policy*, 12(1):127–166.

Coleman, M. R. and Shah-Coltrane, S. (2011). U-STARS∼PLUS: Professional development kit manual. *Arlington, VA: Council for Exceptional Children*.

Costa, A. L. and Kallick, B. (2009). *Habits of Mind across the Curriculum: Practical and Creative Strategies for Teachers*. ASCD.

Darity, W., Castellino, D., Tyson, K., Cobb, C., and McMillen, B. (2001). One Strategy for Closing the Achievement Gap for At-Risk and Ethnic Minority Students. *North Carolina Department of Public Instruction*.

Davis, B., Engberg, J., Epple, D., Sieg, H., and Zimmer, R. (2010). Evaluating the Gifted Program of an Urban School District using a Modified Regression Discontinuity Design. Technical Report w16414, National Bureau of Economic Research, Cambridge, MA.

Dewey, J. (1923). *Democracy and Education: An Introduction to the Philosophy of Education*. macmillan.

Dobbie, W. and Fryer, R. G. (2014). The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools. *American Economic Journal: Applied Economics*, 6(3):58–75.

Donovan, M. S. and Cross, C. T. (2002). *Minority Students in Special and Gifted Education*. National Academies Press, Washington, D.C.

Dottin, E. S. (2009). Professional judgment and dispositions in teacher education. *Teaching and Teacher Education*, 25(1):83–88.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *The American Economic Review*, page 37.

Elhoweris, H., Mutua, K., Alsheikh, N., and Holloway, P. (2005). Effect of Children's Ethnicity on Teachers' Referral and Recommendation Decisions in Gifted and Talented Programs. *Remedial and Special Education*, 26(1):25–31.

Figlio, D. (2005). Names, Expectations and the Black-White Test Score Gap. Technical Report w11195, National Bureau of Economic Research, Cambridge, MA.

Ford, D. Y. (2003). Two other Wrongs Don't Make a Right: Sacrificing the Needs of Diverse Students Does Not Solve Gifted Education's Unresolved Problems. *Journal for the Education of the Gifted*, 26(4):283–291.

Gayle, M. E. (2005). Project Bright IDEA 1: Interest Development Early Abilities. *North Carolina Department of Public Instruction*, page 22.

Gersten, R., Jordan, N. C., and Flojo, J. R. (2005). Early Identification and Interventions for Students With Mathematics Difficulties. *Journal of Learning Disabilities*, 38(4):293–304.

Giessman, J. A., Gambrell, J. L., and Stebbins, M. S. (2013). Minority Performance on the Naglieri Nonverbal Ability Test, Second Edition, Versus the Cognitive Abilities Test, Form 6: One Gifted Program's Experience. *Gifted Child Quarterly*, 57(2):101–109.

Gottfried, M. A. (2009). Excused versus Unexcused: How Student Absences in Elementary School Affect Academic Achievement. *Educational Evaluation and Policy Analysis*, 31(4):392–415.

Grissom, J. A. and Redding, C. (2016). Discretion and Disproportionality: Explaining the

Underrepresentation of High-Achieving Students of Color in Gifted Programs. *AERA Open*, 2(1):233285841562217.

Grissom, J. A., Redding, C., and Bleiberg, J. F. (2019). Money over Merit? Socioeconomic Gaps in Receipt of Gifted Services. *Harvard Educational Review*, 89(3):337–369.

Hemelt, S. W. and Lenard, M. A. (2020). Math acceleration in elementary school: Access and effects on student outcomes. *Economics of Education Review*, page 101921.

Hill, H. C. and Erickson, A. (2019). Using Implementation Fidelity to Aid in Interpreting Program Impacts: A Brief Review. *Educational Researcher*, 48(9):590–598.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86.

Imberman, S. (2021). Ohio's Lost Einsteins: The inequitable outcomes of early high achievers. *Thomas B. Fordham Institute*.

Jackson, C. K. (2018). What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes. *Journal of Political Economy*, 126(5):2072–2107.

Jung, C. G. (1923). *Psychological Types*. Routledge.

Kallick, B. and Zmuda, A. (2017). *Students at the Center: Personalized Learning with Habits of Mind*. ASCD.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4):241–253.

Kraft, M. A., Marinell, W. H., and Shen-Wei Yee, D. (2016). School Organizational Contexts, Teacher Turnover, and Student Achievement: Evidence From Panel Data. *American Educational Research Journal*, 53(5):1411–1449.

Lavy, V. and Goldstein, Y. (2022). Gifted Children Programs' Short and Long-Term Impact: Higher Education, Earnings, and the Knowledge-Economy. *National Bureau of Economic Research*, page 124.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., and 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*.

Liu, J. and Loeb, S. (2019). Engaging Teachers: Measuring the Impact of Teachers on Student Attendance in Secondary School. *Journal of Human Resources*, pages 1216–8430R3.

Lohman, D. F. (2005). Review of Naglieri and Ford (2003): Does the Naglieri Nonverbal Ability Test Identify Equal Proportions of High-Scoring White, Black, and Hispanic Students? *Gifted Child Quarterly*, 49(1):19–28.

McCoach, D. B., Gubbins, E. J., Foreman, J., Rubenstein, L. D., and Rambo-Hernandez, K. E. (2014). Evaluating the Efficacy of Using Predifferentiated and Enriched Mathematics Curricula for Grade 3 Students: A Multisite Cluster-Randomized Trial. *Gifted Child Quarterly*, 58(4):272–286.

Moneta, G. B. and Csikszentmihalyi, M. (1996). The Effect of Perceived Challenges and Skills on the Quality of Subjective Experience. *Journal of Personality*, 64(2):275–310.

Morris, J. E. (2001). African American students and gifted education: The politics of race and culture. *Roeper Review*, 24(2):59–62.

Myers, I. B. (1962). The Myers-Briggs Type Indicator: Manual (1962).

Naglieri, J. A. and Ford, D. Y. (2003). Addressing Underrepresentation of Gifted Minority Children Using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly*, 47(2):155–160.

Naglieri, J. A. and Ford, D. Y. (2005). Increasing Minority Children's Participation in Gifted

Classes Using the NNAT: A Response to Lohman. *Gifted Child Quarterly*, 49(1):29–36.

Nichols, J., Wolfe, M., Thomson, S., and Merritt, D. (2012). *Primary Education Thinking Skills*. Pieces of Learning.

O'Brien, R. (2018). Problem Solving as a Characteristic of Talent Development in the Early Grades. *National Association for Gifted Children*.

Okamoto, Y. and Case, R. (1996). Exploring the Microstructure of Children's Central Conceptual Structures in the Domain of Number. *Monographs of the Society for Research in Child Development*, 61(1-2):27–58.

Olszewski-Kubilius, P. and Clarenbach, J. (2012). Unlocking Emergent Talent:. page 36.

Parks, S. and Black, H. (2012). *Building Thinking Skills: Level 1*. Critical Thinking Company (Bright Minds).

Pearman, F. A. and McGee, E. O. (2022). Anti-Blackness and Racial Disproportionality in Gifted Education. *Exceptional Children*, 88(4):359–380.

Peters, S. J. (2022). Where Does Gifted Education Go From Here: Chaos or Community? *Gifted Child Quarterly*, 66(2):163–168.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In Duncan, G. J. and Murnane, R. J., editors, *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. Russell Sage Foundation.

Reardon, S. F., Fahle, E. M., Kalogrides, D., Podolsky, A., and Zárate, R. C. (2019). Gender Achievement Gaps in U.S. School Districts. *American Educational Research Journal*, page 000283121984382.

Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M., and Kaniskan, R. B. (2011). The Effects of Differentiated Instruction and Enrichment Pedagogy on Reading Achievement in Five Elementary Schools. *American Educational Research Journal*, 48(2):462–501.

Renzulli, J. S. and Reis, S. M. (1985). *The Schoolwide Enrichment Model: A Comprehensive Plan for Educational Excellence*. Creative Learning Press.

Robinson, N. M. (2003). Two Wrongs Do Not Make a Right: Sacrificing the Needs of Gifted Students Does Not Solve Society's Unsolved Problems. *Journal for the Education of the Gifted*, 26(4):251–273.

Shi, Y. (2020). Who benefits from selective education? Evidence from elite boarding school admissions. *Economics of Education Review*, 74:101907.

Shores, K., Kim, H. E., and Still, M. (2020). Categorical Inequality in Black and White: Linking Disproportionality Across Multiple Educational Outcomes. *American Educational Research Journal*, page 000283121990012.

Silver, H. F., Jackson, J. W., and Moirao, D. R. (2011). *Task Rotation: Strategies for Differentiating Activities and Assessments by Learning Style*. ASCD.

Silver, H. F., Strong, R. W., and Perini, M. J. (2007). *The Strategic Teacher: Selecting the Right Research-Based Strategy for Every Lesson*. ASCD.

Snyder, T. D., de Brey, C., and Dillow, S. A. (2020). Digest of Education Statistics, 2018 (54th Edition). *National Center for Education Statistics, Institute of Education Sciences, and U.S. Department of Education*.

Thomson, S. (2009). Effects of Primary Education Thinking Skills on achievement: Mixed-method evaluative study.

Tzur, R. and Watson, M. (2010). Project Bright IDEA 2: Interest Development Early Abilities. *North Carolina Department of Public Instruction*.

VanTassel-Baska, J. and Stambaugh, T. (2007). Overlooked Gems: A National Perspective on Low-Income Promising Learners. Proceedings from the National Leadership Conference on Low-Income Promising Learners (Washington, DC, April 24-25, 2006). *National Association for Gifted*

Children (NJ1).

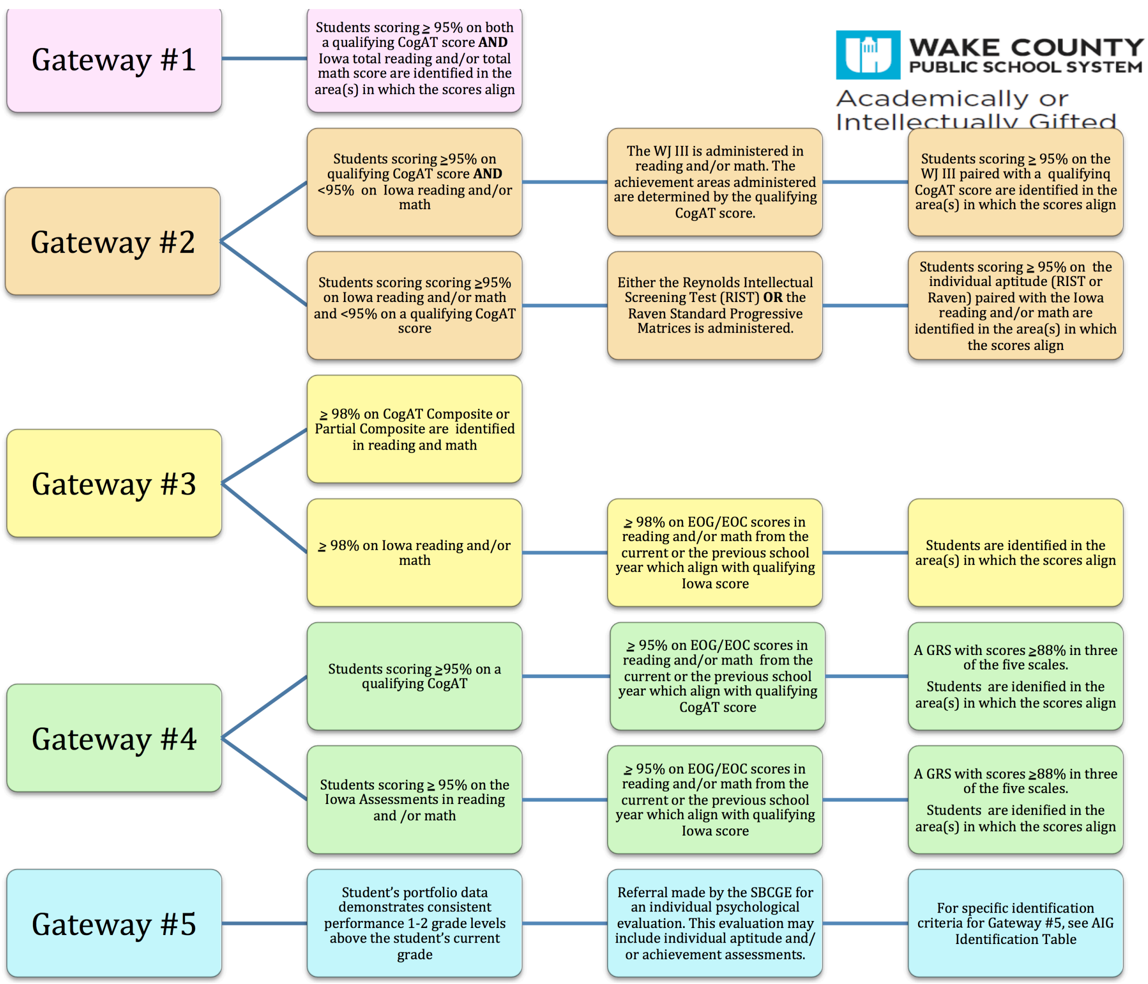
Vygotsky, L. S. (1962). *Thought and Language*. MIT press.

Watson, M., Hargett, M., Tzur, R., and Gayle, M. (2010). Project Bright IDEA 2: Interest Development Early Abilities, a Javits K-2 Nurturing Program Funded by the United States Department of Education, 2004-2009. *North Carolina Department of Public Instruction*, page 5.

Yaluma, C. B. and Tyner, A. (2018). Is There a Gifted Gap? Gifted Education in High-Poverty Schools. *Thomas B. Fordham Institute*.

Yoon, S. Y. and Gentry, M. (2009). Racial and Ethnic Representation in Gifted Programs: Current Status of and Implications for Gifted Asian American Students. *Gifted Child Quarterly*, 53(2):121–136.

A Appendices



Notes: This table displays pre-treatment summary statistics for students in control and treatment groups, as well as for the full sample. Standard errors are clustered at the school level, which is the unit of assignment.

Figure A1: Gifted Identification Gateways, Wake County's 3-year AIG Plan, 2013-2016

Table A1: Student-level Balance Statistics Among Iowa Assessment Takers

Variable	(1) Control	(2) Treatment	(3) Difference
Student characteristics			
Male	0.523	0.554	0.031
	(0.500)	(0.497)	(0.022)
Asian	0.055	0.072	0.016
	(0.229)	(0.258)	(0.034)
Black	0.242	0.196	-0.045
	(0.428)	(0.398)	(0.063)
Hispanic	0.127	0.175	0.047
	(0.334)	(0.380)	(0.039)
White	0.528	0.515	-0.013
	(0.500)	(0.500)	(0.095)
SWD	0.038	0.033	-0.005
	(0.192)	(0.179)	(0.011)
LEP	0.060	0.065	0.006
	(0.237)	(0.248)	(0.020)
Foreign Language at Home	0.149	0.194	0.045
	(0.356)	(0.396)	(0.038)
Observations	919	962	

Notes: This table displays summary statistics for students who qualified to take the Iowa Assessments, which constitutes the second stage of the gifted identification process after they qualify via the CogAT. Columns (1) and (2) display means and standard deviations for the control and treatment group, respectively, and Column (3) displays differences in means. Each row represents a single regression and standard errors are clustered at the school level, which is the unit of assignment.

Table A2: Impacts of Nurturing for Bright Tomorrow (NBT) on Absenteeism

	Log Absences + 1			Any Absences (0/1)			Chronic Absences (0/1)		
	Excused	Unexcused	Total	Excused	Unexcused	Total	Excused	Unexcused	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Cohort 1									
NBT	0.093*	-0.030	0.034	0.034	-0.014	0.009	-0.003	-0.002	-0.002
	(0.041)	(0.050)	(0.027)	(0.017)	(0.014)	(0.006)	(0.003)	(0.005)	(0.009)
Observations	3439	3439	3439	3439	3439	3439	3439	3439	3439
Panel B: Cohort 2									
NBT	0.133**	-0.023	0.050	0.033*	-0.006	-0.007	0.005*	-0.003	0.015
	(0.038)	(0.044)	(0.036)	(0.015)	(0.016)	(0.006)	(0.002)	(0.005)	(0.011)
Observations	3345	3345	3345	3345	3345	3345	3345	3345	3345
Panel C: Cohort 3									
NBT	0.105*	-0.029	0.054	0.011	-0.003	0.001	0.008*	-0.019***	-0.007
	(0.039)	(0.035)	(0.032)	(0.016)	(0.010)	(0.006)	(0.003)	(0.005)	(0.010)
Observations	3233	3233	3233	3233	3233	3233	3233	3233	3233

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays treatment effects of NBT on three categories of absences: (1) excused, (2) unexcused, and (3) total. Each category is specified three separate ways: (1) As the log of absences + 1, (2) as any absences (e.g., non-zero), and (3) as chronic absences, which is commonly defined as 18 or more absences during the school year (i.e., 10% of a 180-day year). Each panel represents one of three cohorts and outcomes are captured at the end of the intervention period—thus, Cohort 1 had three dosage years while Cohort 3 had one year. All models include controls for student demographics, special classifications, and teacher characteristics, and fixed effects for randomization matched pairs. Standard errors are clustered at the school level, which is the level of randomization.

A.1 Teacher Disposition Survey Questions

The implementation team administered the teacher disposition survey (TDS) to treatment and control teachers at the beginning and end of each NBT year (2014-15 through 2016-17). All responses were on a 5-point Likert scale (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree). The original survey included 53 questions. We retained 30 questions that directly asked teachers about their disposition. Items that were reverse-polarized are indicated by "(–)." The full battery of questions was reduced to a single index through principal component analysis.

The survey also included implementation-related questions, such as

1. I look for opportunities to learn more about teaching methods. (+)
2. I look for opportunities to learn more about the subject matters I teach. (+)
3. I look for opportunities to learn more about students' ways of learning. (+)
4. I could foster higher academic results had I taught in a school located in a wealthier neighborhood. (-)
5. To foster creativity among my students I also need to exhibit creativity. (+)
6. Students learn new concepts best when they actively explore problems. (+)
7. I cannot demand students from poor homes to excel academically. (–)
8. My administrators allow me to be an effective instructional leader. (+)
9. I frequently ask my peers for ways to improve my teaching. (+)
10. A teacher must provide a challenging instructional program despite students' difficulties at home. (+)
11. Academic giftedness depends on a teacher's nurturing effort. (+)
12. An effective teacher clearly presents to students what s/he expects them to be able to do. (+)
13. Minority students are more likely to exhibit limited motivation to learn. (–)
14. An effective teacher tailors the curriculum to the students' experience (e.g., omits parts, adds tasks, changes order of topics). (+)
15. In my teaching I tend to be flexible and experiment with the unknown. (+)
16. Regardless of the teacher's intentions and efforts, in every classroom there are several students who cannot reach the intended goals. (–)
17. Students' unique racial background is an important resource in my planning for instruction. (+)
18. I continually involve my students' parents in what we do in class. (+)
19. I seek out opportunities for professional development. (+)

20. An effective, 4-year teacher education program is sufficient for teaching at the K-2 level (hence no further professional development is needed). (–)
21. White students are more likely to exhibit compliance with school norms and regulations than minority students. (–)
22. I get frustrated when asked to teach in ways I was not trained. (–)
23. A teacher should help parents form realistic expectations about their child's giftedness. (–)
24. Gifted students are identified at 3rd grade so as a K-2 teacher I do not have to focus on giftedness. (–)
25. To accomplish my goals I have to consider my students' interests. (+)
26. I use tasks that set up high-level expectations for my gifted students. (+)
27. I use tasks that set up high-level expectations for all my students. (+)
28. I like being a mentor of other teachers. (+)
29. I cannot expect students whose language at home is not standard English to excel academically. (–)
30. In our school, a teacher must devote a substantial amount of energy and time to discipline issues. (–)
31. Academic giftedness is, pretty much, a matter of heredity (nature, not nurture). (–)
32. Students learn well when they can monitor their own work. (+)