Journal of School Psychology 53 (2015) 45–62

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

# Growth and gaps in mathematics achievement of students with and without disabilities on a statewide achievement test☆

Joseph J. Stevens[a,*], Ann C. Schulte[b,1], Stephen N. Elliott[b,2], Joseph F.T. Nese[c,3], Gerald Tindal[c,3]

[a] Department of Educational Methodology, Policy and Leadership, College of Education, 5267 University of Oregon, Eugene, OR 97403-5267, United States
[b] T. Denny Sanford School of Social and Family Dynamics, Arizona State University, PO Box 873701, Tempe, AZ 85287-3701, United States
[c] Behavioral Research and Teaching, 175 Lokey Education, 5262 University of Oregon, Eugene, OR 97403-5262, United States

ARTICLE INFO

*Article history:*
Received 9 February 2014
Received in revised form 3 November 2014
Accepted 4 November 2014
Available online 1 December 2014

*Keywords:*
Exceptional children
Mathematics achievement growth
Achievement gaps
Accountability testing

ABSTRACT

This study estimated mathematics achievement growth trajectories in a statewide sample of 92,045 students with and without disabilities over Grades 3 to 7. Students with disabilities (SWDs) were identified in seven exceptionality categories. Students without disabilities (SWoDs) were categorized as General Education (GE) or Academically/Intellectually Gifted (AIG). Students in all groups showed significant growth that decelerated over grades as well as significant variability in achievement by student group, both at the initial assessment in Grade 3 and in rates of growth over time. Race/ethnicity, gender, parental education, free/reduced lunch status, and English language proficiency were also significant predictors of achievement. Effect size estimates showed substantial year-to-year growth that decreased over grades. Sizeable achievement gaps that were relatively stable over grades were observed between SWoDs and students in specific exceptionality categories. Our study also demonstrated the importance of statistically controlling for variation related to student demographic characteristics. Additional research is needed that expands on these results with the same and additional exceptionality groups.


## 1. Introduction

Foundational skills in mathematics are essential for the development of proficiencies that lead to later achievement and success in school and career (National Early Literacy Panel, 2008), and are a key focus in national efforts to reform education in the United States. However, recent reports of state test score trends have indicated that although students with disabilities (SWDs), have participated in their state assessments in mathematics, the majority of SWDs' performances have not met state proficiency standards (Center on Education Policy, 2009; Thurlow, Altman, Cormier, & Moen, 2008). The purpose of this study was to examine mathematics achievement growth in school children, especially those with disabilities, as a context for understanding the mathematics achievement gap and the implications for policies targeted at closing the achievement gap between students with and without disabilities.

☆ This work was supported by a cooperative agreement with the Institute of Education Sciences, U.S. Department of Education, through grant R32C110004 awarded to the University of Oregon. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the U.S. Department of Education.
* Corresponding author at: Department of Educational Methodology, Policy, and Leadership, University of Oregon, Eugene, OR 97403, United States. Tel.: +1 541 346 2445.

*E-mail addresses:* stevensj@uoregon.edu (J.J. Stevens), ann.schulte@asu.edu (A.C. Schulte), Steve_Elliott@asu.edu (S.N. Elliott), jnese@uoregon.edu (J.F.T. Nese), geraldt@uoregon.edu (G. Tindal).
ACTION EDITOR: Patti Manz.
[1] Tel.: +1 480 965 8372.
[2] Tel.: +1 480 965 8712.
[3] Tel.: +1 541 346 3535.

http://dx.doi.org/10.1016/j.jsp.2014.11.001

### 1.1. The No Child Left Behind Act

The No Child Left Behind Act (NCLB, 2002) represented a fundamental change in educational accountability practices in the United States. It involved the federal government more deeply in state educational practices and created more formal expectations for educational achievement. States were mandated to establish high reading and mathematics proficiency standards at each grade and to evaluate whether students met those standards by conducting annual testing of all students in Grades 3 to 8. Another signature feature of NCLB was its emphasis on the progress of all students, with an explicit goal of closing achievement gaps by 2014 between student groups historically at risk for low achievement relative to the general student population. Progress toward this long-term goal was to be examined yearly through the establishment of adequate yearly progress (AYP) targets stated in terms of the percent of students reaching grade-level proficiency in each grade in mathematics and reading.

Along with the federal emphasis on the progress of all students and closing achievement gaps, NCLB required states to examine AYP by evaluating both the performance of all students relative to the grade-level proficiency standards, and the performance of disaggregated student groups. These disaggregated groups included student race/ethnicity; students living in poverty as indicated by receipt of free or reduced price lunch at school; and students with disabilities. Of the disaggregated groups, the group that has typically displayed the largest achievement gap relative to the general student population has been students with disabilities (SWD). Achievement of SWDs has been a concern for decades (Carlberg & Kavale, 1980; McDonnell, McLaughlin, & Morison, 1997), and currently many states report that over 70% of SWDs perform below proficiency on annual statewide reading and mathematics tests (Center on Education Policy, 2009). Further evidence of the particular difficulties for this group is provided in a study by Eckes and Swando (2009), who found that the most frequent reason for schools' AYP failure was the performance of the SWD group.

Although NCLB (2002) mandates that states establish high and uniform standards for proficiency in mathematics at each grade, there is not a strong empirical basis on which states can draw to establish these proficiency standards or for the expectation that all students will meet them. Knowledge about early development of mathematics skills and abilities is limited (Carlson, Jenkins, Bitterman, & Keller, 2011), with even less information available about the developmental trajectories of mathematics achievement, especially for disaggregated groups including SWDs. This lack of information about developmental trajectories is important because one implication of the establishment of uniform proficiency standards for all students is that groups who start significantly lower in achievement in Grade 3 must attain greater rates of growth to meet grade-level expectations for proficiency.

Adding to the importance of a better understanding of developmental trajectories and the extent of achievement gaps observed at each grade are recent changes to NCLB. The Race To The Top (RTTT) legislation (U.S. Department of Education, 2009) introduced greater flexibility in NCLB requirements for state accountability, including the use of growth models to examine not only current year performance of students but also the degree to which student achievement is progressing toward expectations (Manna & Ryan, 2011). However, we know little about trends in growth for SWDs as a whole or for students in specific exceptionality groups (Wei, Blackorby, & Schiller, 2011; Wei, Lenz, & Blackorby, 2013), and there is no large-scale, published research on mathematics achievement growth of students with specific exceptionalities on the actual state tests used for accountability reporting.

### 1.2. Mathematics achievement and achievement gaps for SWDs

A significant gap in mathematics achievement between students with and without disabilities has been well-documented (Council for Exceptional Children, 2013; Watson & Gable, 2013). For example, on the 2013 National Assessment of Educational Progress (NAEP) mathematics test, in comparison to students without disabilities (SWoDs), much lower percentages of SWDs demonstrated performance at or above “proficient” in Grade 4 (18% vs. 45%) and Grade 8 (9% vs. 39%; U.S. Department of Education, 2013). This gap is further documented on a variety of academic achievement tests. Specifically, researchers have consistently noted the lower mathematics performance of SWDs and differences in skills and abilities that may persist from early learning through later grades (Carlson et al., 2011; Chatterji, 2005; Denton & West, 2002; Lee & Burkam, 2002; LoGerfo, Nichols, & Reardon, 2006; Morgan, Farkas, & Wu, 2009; Princiotta, Flanagan, & Germino-Hausken, 2006; Shin, Davison, Long, Chan, & Heistad, 2013; Wei et al., 2013).

Although NCLB (2002) treats SWDs as one undifferentiated group, specific exceptionality categories represent very different kinds of learners whose average performance may differ significantly (Geary, Hoard, Nugent, & Bailey, 2012; Morgan, Farkas, & Wu, 2011; Wei et al., 2013). For example, in a meta-analysis of characteristics of different exceptionality groups, Sabornie, Cullinan, Osborne, and Brock (2005) found that compared to students with emotional/behavioral disabilities, students with mild intellectual disabilities, on average, scored two thirds of a standard deviation lower in academic achievement. Similarly, Morgan et al. (2011) found that, in fifth grade, students with speech-language impairments were approximately one half standard deviation below SWoDs, whereas students with learning disabilities were more than one standard deviation below SWoDs. So, although NCLB focuses on the achievement gap for SWDs as a whole, it is important to investigate and document the extent to which the achievement gap differs for specific exceptionality groups.

Central to NCLB and RTTT policy is the goal of closing or eliminating the achievement gap as students progress across grades. However, longitudinal studies of achievement gaps for SWDs report varying results. A number of studies have found increases in achievement gaps over time using district, state, and federal mathematics and reading assessments (Geary et al., 2012; Judge & Watson, 2011; Morgan et al., 2011; Wei et al., 2013). Less frequently, studies have reported stable achievement gaps over time (Jordan, Kaplan, & Hanich, 2002; Scarborough, 1998; Shaywitz et al., 1995). Finally, a number of researchers have reported decreases in the achievement gap, again across a variety of mathematics and reading assessments including district, state, and federal tests (Bast & Reitsma, 1998; Ding, Davison, & Petersen, 2005; Galindo, 2010; Han, 2008; Jordan et al., 2002; Protopapas, Sideridis, Mouzaki, & Simos, 2011; Scarborough & Parker, 2003; Tate, 1997).

### 1.3. Mathematics achievement growth for SWDs, SWoDs, and gifted students

In contrast to the varying results for achievement gaps, most investigations of student mathematics achievement growth across grades have generally found consistent results. In studies that have examined mathematics growth of all students, mathematics achievement is usually characterized by a curvilinear function with relatively high rates of growth in the early grades that progressively decelerate in later grades (Bloom, Hill, Black, & Lipsey, 2008; Carlson et al., 2011; Ding et al., 2005; Lee, 2010; Morgan et al., 2011; Shin et al., 2013; Wei et al., 2013). For instance, Lee (2010) used multiple sources of national assessment data, including long-term NAEP trend data, the Early Childhood Longitudinal Study—Kindergarten (ECLS-K), and norms from two standardized achievement tests to examine mathematics growth in the United States over three decades. Lee characterized typical mathematics achievement growth as consisting of an overall achievement gain of 6 to 7 standard deviations from kindergarten entry to high school exit. Typical gains were one standard deviation per grade in the primary grades, a half standard deviation per grade in middle school, and a quarter of a standard deviation per grade in high school. Bloom et al. (2008) reached similar conclusions about the general pattern of decelerating mathematics achievement growth by examining average annual gains in mathematics achievement from Kindergarten to Grade 12 on six nationally normed tests. This general finding of decelerating growth in mathematics achievement is widespread and has been attributed to a number of factors including decreasing rate of growth in children's cognitive capacity with age, increasing complexity of mathematics content (Lee, 2010), changes in the structure of the content assessed, or differences in the assessments and instruments used to document performance and progress.

Until recently, relatively few studies have examined achievement growth for specific groups of exceptional children including SWD and Academically/Intellectually Gifted (AIG) students. Some investigators (e.g., Shin et al., 2013) have examined mathematics achievement growth trajectories using a dichotomous categorization of SWDs vs. SWoDs. A number of other investigators have examined one or two specific exceptionalities—most commonly speech-language impairment, specific learning disabilities (LD), or both conditions. For example, Judge and Watson (2011) investigated mathematics achievement growth of students with specific learning disabilities using data from ECLS-K over Grades K to 5. Results showed that lower levels of mathematics achievement were already present at kindergarten entry for students identified as learning disabled and these students had slower growth than SWoDs. Morgan et al. (2011), in a study of ECLS-K participants, found that children identified with either speech-language impairments or learning disabilities performed significantly lower than children without disabilities at kindergarten entry but only children with LD showed significantly slower mathematics growth than SWoDs across the elementary school grades.

Investigators rarely have examined achievement growth in multiple disability categories (e.g., Wei et al., 2011, 2013), but when they have, considerable heterogeneity appears in intercept and some heterogeneity in slope of growth trajectories for different exceptionalities. Wei et al. (2013), using the nationally representative Special Education Elementary Longitudinal Study (SEELS), estimated mathematics achievement growth trajectories for students in 11 specific disability categories from age 7 to 17 using quadratic growth models. Wei et al. (2013) did not provide direct comparisons of exceptional children's growth to SWoDs as those students were not included in the SEELS sample. Instead, students with learning disabilities, who constituted the largest proportion of the sample, were the reference group in statistical analyses. Overall, mathematics growth for all SWDs followed a pattern similar to that observed in previous studies of SWoDs (e.g., Lee, 2010) with curvilinear growth that progressively decelerated through high school. The SEELS outcome measures assessed applied mathematics and mathematics calculation. On the applied mathematics measure at the midpoint of the age range, Wei et al. (2013) found that average performance (intercepts) of students in all disability categories was significantly lower than students with learning disabilities (the reference group) with the exception of students with speech impairments, emotional disturbance, or visual impairments. The ranking of disability groups from highest to lowest average performance was speech impairments, visual impairments, emotional disturbances, learning disabilities, other health impairments, orthopedic impairments, hearing impairments, traumatic brain injury, autism, intellectual disability, and multiple disabilities. In contrast, there were no statistically significant differences between students with learning disabilities and students in each of the other disability categories on slope or acceleration of growth trajectories. On the mathematics calculation outcome measure, Wei et al. found that, on average, students with autism, intellectual disabilities, traumatic brain injury, or multiple disabilities had significantly lower scores and students with speech impairments had significantly higher scores in comparison to students with learning disabilities. Wei et al. found no statistically significant differences in slope or acceleration of calculation scores between students with learning disabilities and students in each of the other disability categories, with the exception of a significantly slower growth rate for students with autism and a significantly faster deceleration for students with speech impairments.

There is also little published research on the academic growth of AIG students. In the research reported above on the achievement gap, in NAEP achievement reports, and in almost all state accountability systems, AIG students are not differentiated from general education (GE) students and together they form the student group typically referred to as SWoDs. Although historically there have been some studies of AIG students' development, including longitudinal studies (see Lubinski, Webb, Morelock, & Benbow, 2001), with rare exceptions, this research has been based on case studies and small samples and has not examined these students' performance on academic achievement tests. The Association for the Gifted of the Council on Exceptional Children (CEC-TAG, n.d.) has recently argued for the separate evaluation of the achievement performance and academic growth of AIG students as part of NCLB requirements. Subotnik, Olszewski-Kubilius, and Worrell (2011) described reasons why giftedness is often excluded from consideration in educational policy, including the assumption that AIG students will "do OK on their own" (p. 8). They argued for the importance of clearly identifying and studying gifted children but also noted the unique methodological challenges posed by this group, such as the need for longitudinal study to ameliorate regression to the mean artifacts (Lohman & Korb, 2006).

In one of the few studies of achievement growth in AIG students, Ma (2005) used data from the Longitudinal Study of American Youth over Grades 7 to 12 and found that provision of an early, accelerated mathematics curriculum resulted in little improvement in mathematics growth among AIG students, small improvements among honors students, and larger improvements among GE students. Ma provided evidence that ceiling effects, a common problem in the study of academic growth in AIG students, did not affect the outcome measure. In another study, Rambo-Hernandez and McCoach (2014) found that, on the Measures of Academic Progress (MAP) reading test (Northwest Evaluation Association, 2011), average students' initial performance in Grade 3 was lower than AIG students, but average students had greater growth during the school year and through to the fall of Grade 6. AIG students, however, grew more quickly than average students over the summer and maintained the same growth rate throughout the calendar year. In sum, despite recommendations for disaggregation and study of the academic growth of AIG children, there are a number of methodological challenges and few extant longitudinal studies.

### *1.4. Relations of mathematics achievement and growth to student socio-demographic characteristics*

Examination of mathematics achievement gap differences has largely focused on groups defined by NCLB (i.e., economic disadvantage, language proficiency, race/ethnicity, and SWDs). Substantial evidence indicates that economically disadvantaged students start school behind their more advantaged peers, have fewer resources in the home, and fewer opportunities for mathematics learning (Davison, Seok Seo, Davenport, Butterbaugh, & Davison, 2004; Denton & West, 2002; Jordan et al., 2002; Lee & Burkam, 2002). Evidence also exists indicating that these differences persist throughout the school years (Chatterji, 2005; McCoach, O'Connell, Reis, & Levitt, 2006; Morgan et al., 2009; Wright & Li, 2008).

Investigators have also found significant associations between student achievement and both gender and parent education level (Holman, 1995; Hyde, Fennema, & Lamon, 1998; Phillips, Brooks-Gunn, Duncan, Klebanov, & Crane, 1998). Well established differences have been found in mathematics achievement by student race/ethnicity group on national achievement tests as well as other academic, behavioral, and career outcomes (Blackorby et al., 2005; Blair & Scott, 2002; Denton & West, 2002; Hemphill, Vanneman, & Rahman, 2011; Jordan et al., 2002; Reardon & Galindo, 2009; Vanneman, Hamilton, Anderson, & Rahman, 2009). The importance of taking student socio-demographic characteristics into account in examining achievement growth trajectories has also been demonstrated in several studies. For example, Kieffer (2011) found that differences in growth trajectories between native English speakers and English language learners (ELLs) decreased when child- and school-level socioeconomic status (SES) measures were taken into account with the achievement gap narrowing over time and ELLs growing more rapidly than native English language peers by eighth grade.

Relatively few investigators of mathematics achievement have explored the impact of socio-demographic variables on growth or change in achievement gaps for SWDs. In a study of mathematics learning disabilities, Judge and Watson (2011) found that students participating in ECLS-K had significantly slower achievement growth over Grades K to 5 if they were female, African American, Hispanic, or lower SES. In another analysis of ECLS-K respondents, Morgan et al. (2009) found that non-White, female, lower SES students identified with an individualized education plan scored significantly lower on initial status and had slower mathematics growth rates over Grades 1 to 5 than peers. In a larger study of ECLS-K participants, Morgan et al. (2011) reported that children identified with speech impairments or learning disabilities who were from lower socioeconomic status families or were African American had lower levels of mathematics achievement and lagged increasingly behind in their acquisition of mathematics skills over time.

As described earlier, using SEELS data, Wei et al. (2013) analyzed mathematics growth trajectories for students in 11 federal disability categories. Wei et al. also included student gender, race, and socioeconomic status as additional predictors in their analyses. They found significantly lower mathematics achievement for SWDs who were female, Black, and lower SES in comparison to White, male, average SES students with learning disabilities. They also observed that achievement gaps were stable over time for these groups but widened for Hispanic students with disabilities on mathematics calculation.

### *1.5. Purpose and research questions*

Although there is an informative and developing literature on the mathematics growth of SWDs, there are important gaps and limitations in the literature. There are only three studies (Carlson et al., 2011; Wei et al., 2011, 2013) that examined growth for more than two specific student exceptionality groups. To our knowledge, there are no large-scale, published studies that examined mathematics growth of specific student exceptionality groups using a state accountability test used for high stakes reporting under NCLB. Other limitations of some previous research are the failure to include specifically identifiable SWoD reference groups as comparisons in analysis, the lack of interpretative benchmarks for growth (as recommended by Bloom et al., 2008), and only descriptive evidence or visual inspection of achievement gap differences.

The purpose of the present study was to add to the literature by examining mathematics achievement growth on a statewide achievement test for students in specific exceptionality categories in Grades 3 to 7 using descriptive methods, multilevel longitudinal models, clearly identified reference groups, and empirical, effect size (ES) benchmarks of growth. We also sought to refine our estimates of growth trajectories by including student socio-demographic characteristics in the multilevel analyses. Interpretive growth benchmarks were provided by describing the ES of students' mathematics achievement from year-to-year for all student groups and the ES of achievement gaps of specific student exceptionality groups in comparison to GE students and in comparison to SWoDs (GE and AIG students together) as is done in accountability reporting.

We addressed five fundamental questions about academic growth in mathematics: (a) What is the developmental progress in mathematics achievement for GE students, AIG students, and students in specific exceptionality groups on a statewide achievement test? (b) Do rates of growth change over grades? (c) Do rates of growth change after controlling for student socio-demographic characteristics? (d) What is the ES of growth from year-to-year for specific exceptionality groups? and (e) What is the ES of the achievement gap between students in specific exceptionality categories and GE students or SWoDs as a whole and does the magnitude of gaps change over grades?

## 2. Method

### 2.1. Sample

The original sample for this study consisted of all students in North Carolina who were in the third grade in 2000–2001, had not been retained in grade from the previous year, and were present in the state's achievement testing database ($N = 103{,}123$; see first column of Table 1 labeled Total sample). We selected North Carolina as the focal state for this analysis for several reasons. The state's educational accountability system has employed reading and mathematics tests with developmental scales spanning grades three through eight since 1993 (North Carolina Department of Public Instruction, 1996). These developmental scales have a strong psychometric foundation (Thissen, Pommerich, Billeaud, & Williams, 1995; Williams, Pommerich, & Thissen, 1998) and are well described in the tests' technical manuals (North Carolina Department of Public Instruction, 1996, 2006). In addition, unlike many states, North Carolina has routinely included students with disabilities in their annual accountability testing program well before federal mandates to do so (Schulte, Villwock, Whichard, & Stallings, 2001).

Based on the 2000 United States Census (U. S. Census Bureau, 2014), which is the closest in time to the data reported here, North Carolina was similar to the U.S. general population in terms of percentage of Whites (72.1 versus 75.1) and American Indians (1.2 versus 0.9), but the state had a higher percentage of the population that was Black (21.6 versus 12.3) and smaller percentages who were Asian (1.4 versus 3.6) or Hispanic (4.7 versus 12.5). Students in North Carolina scored one quarter of a standard deviation above the national average at fourth grade on NAEP mathematics in 2003, the closest testing period to the year this longitudinal sample was in fourth grade (U. S. Department of Education, 2013). The percent of students with disabilities as a percent of public school enrollment in North Carolina (13.4) closely approximated the national percent in 2001 (13.3; U.S. Department of Education, 2002).

To create the analytic sample, we systematically excluded a number of individuals (see second column of Table 1 labeled Analysis sample) if they met one or more of the following criteria: (a) did not follow the typical grade level sequence from Grades 3 to 7, primarily due to grade retention ($N = 8315$, 8.1%); (b) never participated in the large scale mathematics test in any grade from 3 to 7 ($N = 1729$, 1.7%); or (c) were missing wave one, third-grade demographic information, including codes for ethnicity ($N = 14$, <0.1%), gender ($N = 11$, <0.1%), parental education ($N = 1206$, 1.2%), or exceptionality ($N = 314$, <1%). We also excluded 30 students coded as "other" ethnicity in 2001 because the category was dropped by North Carolina in subsequent years. Retained students are likely to be systematically different than other students and might have biased our estimation of "typical" growth trajectories. Although we could have used imputation or other statistical methods to replace missing demographic information, research has shown that listwise deletion is an acceptable strategy with very small percentages (<2%) of missing data or when missingness occurs on predictor variables, as is the case here (Enders, 2010). Finally, in order to ensure stable statistical estimation and power to examine growth for students in specific exceptionality categories, we also dropped from the analytic sample several categories with a sample size of 100 or less in third grade (multiple disabilities, orthopedic impairment, traumatic brain injury, and visual impairment, $N = 321$, <1%).

After students meeting one or more of these exclusion criteria had been eliminated, the analytic sample consisted of 92,045 students (89% of the students in the state system data file and 91% of students tested). Table 1 shows the composition of the analytic and total sample with respect to SWoD, SWD, and specific exceptionality groups. We used $z$-tests of the difference between

**Table 1**
Student disability group by sample at wave 1.

| Student group | Total sample | | Analysis sample | | $h$ |
|---|---|---|---|---|---|
| | $N$ | % | $N$ | % | |
| Students without disabilities | 88,429 | 85.7 | 81,179 | 88.2 | .074 |
| General education | 81,478 | 79.0 | 74,333 | 80.8 | .045 |
| Academic/intell. gifted | 6951 | 6.7 | 6846 | 7.4 | .027 |
| Students with disabilities | 14,694 | 14.3 | 10,866 | 11.8 | .074 |
| Autism | 332 | 0.3 | 141 | 0.2 | .020 |
| Emotional disturbance | 829 | 0.8 | 634 | 0.7 | .012 |
| Hearing impairment | 170 | 0.2 | 131 | 0.1 | .026 |
| Mild intellectual disability | 2017 | 2.0 | 1244 | 1.4 | .047 |
| Multiple disability | 100 | 0.1 | – | | – |
| Orthopedic impairment | 80 | 0.1 | – | – | |
| Other health impairment | 1502 | 1.5 | 1171 | 1.3 | .017 |
| Specific learning disability | 6377 | 6.2 | 5221 | 5.7 | .021 |
| Speech-language impairment | 2660 | 2.6 | 2324 | 2.5 | .006 |
| Traumatic brain injury | 29 | <0.1 | – | | – |
| Visual impairment | 59 | <0.1 | – | | – |
| Unidentified | 314 | 0.3 | – | | – |
| Total sample size | 103,123 | | 92,045 | | |

*Note*. Students without disabilities group is composed of the combination of general education and academically/intellectually gifted students.

proportions to compare the two samples. Because of our very large sample sizes, any difference of .1% or larger between the two samples was statistically significant ($p < .05$). To more fully evaluate these differences we also computed Cohen's $h$, a measure of effect size for proportions. Cohen provides rules of thumb for $h$ (Cohen, 1988, pp. 184–85), interpreting effect sizes of .20 as *small*, about .50 as *medium*, and .80 and greater as *large*. As can be seen in Table 1, all differences between samples were very small ranging from an $h$ of .006 for speech-language impairment to .074 for SWD overall.

We also evaluated the samples on their demographic characteristics including differences between the SWoD and SWD groups (see Table 2). As with the comparisons in Table 1, we used $z$-tests for proportions and Cohen's $h$ to compare the total sample to the analytic sample and the SWoD to the SWD groups within the analytic sample. Although all but two comparisons (Asian and Multi-racial) between the Total and Analysis samples were statistically significant, the magnitudes of these differences were very small for all 16 comparisons in Table 2 ranging from an $h$ of 0 to an $h$ of 0.047. Comparisons of the demographic characteristics of the SWoD and SWD groups, however, showed more variation. Although many comparisons between the groups were also very small in magnitude, there were also several larger differences in the composition of the SWD group including 20% fewer female students ($h = 0.41$), 12% higher participation in free/reduced lunch ($h = 0.24$), and more students with lower parental educational levels (e.g., 10% more students with parents who had less than a High School education, $h = .28$).

We forward matched the cohort of students who were present in the database in 2000–2001 in Grade 3 to all succeeding years through Grade 7 (2004–05). The forward matching procedure tracked all students in the system in 2000–2001 but did not add new students entering the system in later years (e.g., new Grade 4 students in 2001–2002 and new Grade 5 students in 2002–2003). We used forward matching to identify a clearly defined cohort of students and avoid a shifting student composition over the grades studied. We tracked the cohort only through seventh grade because the state introduced a third edition of the mathematics test in 2005–2006 that would have confounded estimates of mathematics growth with score changes due to test edition. Of the 92,045 students in the analytic sample, 80.9% had mathematics scores in all 5 years, 5.1% had scores in 4 years, 4.3% had scores in 3 years, 3.8% had scores in 2 years, and 6.0% had only one mathematics score during the 5-year study period. Achievement data for a student could be missing in a year because the student took an alternate assessment, began attending a private school that did not participate in the testing program, or moved out of state.

### *2.2. Measures*

For all analyses reported, the outcome measure was student developmental scale score on the standardized, second edition North Carolina End of Grade (EOG) Mathematics Tests. Each grade level test consisted of 80 multiple-choice items intended to measure the four strands in the state mathematics curriculum: (a) number sense, numeration, and numerical operations; (b) spatial sense, measurement, and geometry; (c) patterns, relationships, and functions; and (d) data, probability, and statistics (North Carolina Department of Public Instruction, 2006). Although the content specifications were based on strands, the scores reported are total scores across all strands. For reference, the central tendency and variability of these total scores across all North Carolina students are described by the values in the first two rows of Table 3. North Carolina mathematics test scores reported here are developmental scale scores based on vertical linking. The linking study used to establish the developmental scale employed a common items design in which adjacent grades were forward linked using a three-parameter logistic item response theory model, 12-item linking forms between adjacent grades, and an algorithm described by Thissen and Orlando (2001).

**Table 2**
Student demographic characteristics by sample and group at wave 1.

| Characteristic | Total sample | | Analysis sample | | | SWoD | | SWD | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | % | *N* | % | *h* | *N* | % | *N* | % | *h* |
| Female | 50,463 | 48.9 | 46,364 | 50.4 | .030 | 42,819 | 52.7 | 3545 | 32.6 | .409 |
| American Indian | 1549 | 1.9 | 1353 | 1.5 | .031 | 1166 | 1.4 | 187 | 1.7 | .024 |
| Asian | 1958 | 1.9 | 1791 | 1.9 | .000 | 1720 | 2.1 | 71 | 0.7 | .123 |
| Black | 31,190 | 30.2 | 26,096 | 28.4 | .040 | 22,640 | 27.9 | 3456 | 31.8 | .085 |
| Hispanic | 5555 | 5.4 | 4555 | 4.9 | .023 | 4233 | 5.2 | 322 | 3.0 | .112 |
| Multi-racial | 1818 | 1.8 | 1639 | 1.8 | .000 | 1458 | 1.8 | 181 | 1.7 | .008 |
| White | 61,005 | 59.2 | 56,611 | 61.5 | .047 | 49,962 | 61.5 | 6649 | 61.2 | .006 |
| Unidentified | 17 | <0.1 | – | | – | | | – | | |
| Limited English | 3553 | 3.4 | 2724 | 3.0 | .023 | 2555 | 3.1 | 169 | 1.6 | .100 |
| Title I student | 4661 | 4.5 | 3882 | 4.2 | .015 | 3328 | 4.1 | 554 | 5.1 | .048 |
| Free/reduced lunch | 40,189 | 39.1 | 37,266 | 40.5 | .029 | 1744 | 39.1 | 5522 | 50.8 | .236 |
| Parental education | | | | | | | | | | |
| <High school | 12,158 | 11.8 | 9532 | 10.4 | .045 | 7482 | 9.2 | 2050 | 18.9 | .283 |
| High school | 47,247 | 45.8 | 41,756 | 45.4 | .008 | 36,346 | 44.8 | 5410 | 49.8 | .100 |
| High school + | 4148 | 4.0 | 3848 | 4.2 | .010 | 3470 | 4.3 | 378 | 3.5 | .041 |
| Com. college graduate | 13,368 | 13.0 | 12,599 | 13.7 | .021 | 11,437 | 14.1 | 1162 | 10.7 | .103 |
| Trade/bus. sch. graduate | 20,771 | 20.1 | 20,179 | 21.9 | .044 | 18,628 | 22.9 | 1551 | 14.3 | .222 |
| College graduate | 4225 | 4.1 | 4131 | 4.5 | .020 | 3816 | 4.7 | 315 | 2.9 | .095 |
| Total sample size | 103,123 | | 92,045 | | 81,179 | | | 10,866 | | |

**Table 3**
Mathematics scale score means and standard deviations by student group ($N = 92{,}045$).

| | Grade | | | | |
|---|---|---|---|---|---|
| Student group | 3 | 4 | 5 | 6 | 7 |
| All students | 251.44<br>(7.50) | 257.39<br>(8.21) | 263.22<br>(8.68) | 267.15<br>(9.18) | 270.15<br>(10.88) |
| General education | 251.17<br>(6.91) | 257.04<br>(7.61) | 263.06<br>(7.91) | 267.16<br>(8.34) | 269.98<br>(9.94) |
| Acad./intell. gifted | 260.80<br>(5.55) | 267.64<br>(6.30) | 273.76<br>(6.12) | 277.85<br>(6.40) | 283.43<br>(8.90) |
| Autism | 246.30<br>(7.76) | 252.39<br>(8.61) | 256.43<br>(10.59) | 260.49<br>(10.39) | 263.30<br>(12.49) |
| Emotional disturbance | 243.98<br>(7.06) | 249.91<br>(7.13) | 254.19<br>(8.59) | 256.53<br>(8.47) | 257.77<br>(8.70) |
| Hearing impairment | 247.79<br>(7.10) | 253.28<br>(7.46) | 259.65<br>(8.27) | 262.16<br>(9.01) | 264.21<br>(10.02) |
| Mild intellectual disability | 237.56<br>(4.79) | 243.25<br>(4.77) | 246.68<br>(5.26) | 249.61<br>(5.46) | 252.24<br>(5.85) |
| Other health impairment | 246.27<br>(7.00) | 251.49<br>(7.50) | 255.79<br>(8.27) | 258.67<br>(8.67) | 261.07<br>(9.48) |
| Spec. learning disability | 247.03<br>(6.80) | 252.23<br>(7.08) | 257.22<br>(8.07) | 260.28<br>(8.44) | 262.44<br>(9.11) |
| Speech-language impair. | 249.74<br>(7.46) | 255.85<br>(8.16) | 261.80<br>(8.80) | 265.95<br>(9.28) | 268.62<br>(10.60) |

A technical manual is available that provides reliability and validity data for the EOG Mathematics Tests, as well as information about scaling and standard setting (North Carolina Department of Public Instruction, 2006). Internal consistency estimates for the mathematics total scale score were above .90 for Grades 3 to 8 and remained above .90 when examined separately by gender, ethnicity, disability status, or limited English proficiency status. Content validity evidence provided in the technical manual included teachers' ratings of the test content, with 85% agreeing to a superior or high degree that the content reflected the goals and objectives of the grade level mathematics curriculum. In terms of criterion-related validity evidence, teachers' ratings of student expected grades in mathematics correlated from .56 to .69 with students' scale scores, with correlations reported separately by grade. In addition, teachers' ratings of student grade-level mastery of mathematics rated on a four-point scale (with descriptors ranging from insufficient to superior mastery) correlated from .61 to .65 with students' scale scores. None of the correlations of scale scores with teacher ratings were reported separately by disability status.

The technical manual reported generally low correlations among EOG scale scores and variables external to the test such as gender, limited English proficiency, and student disability status. The majority of the correlations between scale scores and gender or limited English proficient were less than $\pm$.10, and most of the correlations between scale scores and disability status were less than $\pm$.30. None of these relations approached the size of coefficients for concurrent validity. These findings generalized across all content areas and test forms, and across Grades 3 to 8. Although not reported in the technical manual, evidence of concurrent and predictive validity for the EOG Mathematics Tests was provided in several additional studies. Williams, Rosa, McLeod, Thissen, and Sanford (1998) linked a shortened version of the EOG Mathematics Tests (1st edition) with NAEP mathematics scores. In a summary of this study, Thissen (2007) reported that the EOG Mathematics Tests–NAEP correlation was .73. Renaissance Learning (2008) reported the correlation between the third-grade EOG Mathematics Test (2nd edition) and the STAR Mathematics test was .73. In a more recent concurrent validity study (Northwest Evaluation Association, 2014), the NWEA MAP mathematics test correlated from .81 to .86 over Grades 3 to 8 with the EOG Mathematics Tests (3rd edition) at corresponding grades. None of these studies reported concurrent and predictive validity results separately for SWDs and SWoDs.

### *2.3. Procedures*

North Carolina began its current large-scale assessment and accountability program in reading and mathematics in 1993. Test data are housed for research purposes at the North Carolina Education Research Data Center (NCERDC) at Duke University. The testing program and procedures used to build the longitudinal dataset used are described next.

#### *2.3.1. Test administration*

The North Carolina EOG Mathematics Tests are part of the North Carolina Department of Public Instruction's school accountability program and are administered to students in May of each year, typically in the general education classroom by general education teachers and proctors. Generally, state testing guidelines require students to be included in testing, but students may be exempted from testing or take an alternate assessment for a variety of reasons, including limited English proficiency or determination by an IEP team that a student with a disability should not participate in testing (North Carolina Department of Public Instruction, 2006). For the 2000–2001 school year, the participation rate for the EOG Mathematics Test for students in the third grade was 98% overall and 81.6% for SWDs. Within the special education population, the participation rate by disability varied from 0% for students with moderate to severe intellectual disabilities to 91% for students with learning disabilities.

#### 2.3.2. Determination of student groups

Unlike many states, North Carolina explicitly identifies a group of students labeled as "Academically/Intellectually Gifted" (AIG). We took advantage of this opportunity to study growth of this group and also to compare the growth of student groups not only to GE students but also to AIG students by including this group in our analyses. Exceptional children were identified as SWDs and in need of special education services at the school level, and in the third grade, students were identified in 1 of 17 categories. These classifications were used as the basis for the exceptionality groups used in this study. Two of these categories were for students who were not receiving special education services (a) GE students not identified as exceptional and (b) AIG students. The remaining 15 categories paralleled the disability categories in the Individuals with Disabilities Education Improvement Act (IDEIA, 2004), although the terms sometimes varied somewhat from the IDEIA designations (e.g., "Behaviorally–Emotionally Handicapped" for IDEIA's "Emotional Disturbance").

Prior to data analysis, we recoded the state's exceptionality classifications to match the IDEIA designations with two exceptions. The IDEIA designations do not include an academically gifted category. Because we were interested in academic growth for these students, we analyzed this group separately except in some of the reporting of achievement gap differences where we conformed to NCLB practice of comparing SWD performance with all students without disabilities (i.e., GE students with AIG students together). The second exception to IDEIA labels involves the category of intellectual disability. In North Carolina, this category is separated into mild, moderate, and severe intellectual disabilities; however, only the "mild" designation included the subset of students who had participated in the primary state mathematics assessment reported here. Because this group is not directly comparable to the IDEIA category, we maintained the North Carolina designation of students with a "mild intellectual disability."

#### 2.3.3. Longitudinal database construction

The study dataset was constructed from multiple annual student electronic files available from the NCERDC. At each school, there is one EOG test record for each student who was a member of that school at the time of end of year mathematics and reading testing, even if the student was absent or exempt from testing. Also included in the EOG record is student demographic information and disability classification coded by school personnel at the time of testing and a unique identifier for each student, added by NCERDC, used to match student records across years. To create the longitudinal records, we conducted additional data quality checks on the EOG test record files for the years 2001 to 2005 and then merged annual files by student identification number to create the longitudinal dataset.

### 2.4. Analytic methods

We used multiple analytic approaches to describe and estimate the growth in mathematics achievement of students with and without disabilities. First, we applied descriptive methods to report mean change by year for all students by student group. Second, similar to most studies published in this area, we modeled student growth using multilevel longitudinal analyses (Raudenbush & Bryk, 2002). The multilevel analyses were completed using HLM 7.0 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011), full maximum likelihood estimation, and specification of model parameters as random effects. We did not include a third school level in the analyses because our interest was in student growth rather than school effects, and we wished to avoid attrition of students who changed schools. In the two-level HLM models we used, time was centered at the first testing occasion (Grade 3), and we included data from any available time point for a student even when data from other time points were missing.

We first applied unconditional growth models to serve as a baseline for comparison to succeeding models. The next modeling step applied conditional multilevel growth models to examine model results for each student exceptionality category. The last step in our analyses was to add student characteristics and demographic variables as additional predictors of mathematics achievement. The conditional models included a level-1 model that specified student mathematics scores predicted by a quadratic function of time of measurement and a level-2 model composed of the prediction of level-1 model parameters as a function of student exceptionality categories and demographic characteristics. We used a quadratic model based on previous findings in the literature as well as inspection of the data and statistical testing of alternative growth functions.[4] The initial level-1 model can be described as follows:

$$(Y_{ti}) = \pi_{0i} + \pi_{1i}(\text{time}) + \pi_{2i}(\text{time squared}) + r_{ti} \tag{1}$$

where $Y$ is the mathematics scale score for student $i$ at time $t$ and $\pi_{0i}$ is the initial status or intercept for student $i$ at time 0, $\pi_{1i}$ is the linear rate of change, $\pi_{2i}$ is the quadratic curvature and represents the acceleration or deceleration in each student's growth trajectory and $r_{ti}$ is the residual for each student.

At level-2, the level-1 parameters were modeled using student characteristics. All dichotomous predictors were uncentered, but parental education, a continuous variable coded into ordinal categories, was centered at the grand mean. With this approach, coefficients for dichotomous predictors represented the effect for the group coded one and the coefficient for parental education

[4] We also examined four additional fractional polynomial models representing other degrees of curvature between a linear and a quadratic model and a cubic model. Although a cubic function was statistically significant, it accounted for less than one-tenth of 1% of additional variance explained. Based on inspection of fit indices (i.e., Akaike information criterion, Bayes information criterion, and deviance), other findings in the literature, and visual inspection, a quadratic model was deemed the most appropriate and parsimonious model. Details on these other analyses are available from the first author.

represented the effect for a student with average parental education. The level-2 equations for the mathematics initial status and growth rate parameters were as follows:

$$Initial\ \ Status,\ \ \pi_{0i} = \beta_{00} + \sum \beta_{0k}(Predictor_k) + u_{0i} \tag{2}$$

$$Rate\ \ of\ \ Change,\ \ \pi_{li} = \beta_{10} + \sum \beta_{1k}(Predictor_k) + u_{1i} \tag{3}$$

$$Curvature,\ \ \pi_{2i} = \beta_{20} + \sum \beta_{2k}(Predictor_k) + u_{2i} \tag{4}$$

where $\beta_{00}$ is the mathematics score intercept at Grade 3 for all students, each $\beta_{0k}$ represents the average partial regression coefficient relating the predictor of interest to student's initial status, and $u_{0i}$ is the residual between the fitted predictor value for each student and the student's observed Grade 3 mathematics score. For each rate of change parameter (i.e., $\pi_{1i}$ and $\pi_{2i}$), each individual's rate of change, $\pi_{pi}$, was modeled as a function of the average mathematics rate of change, $\beta_{p0}$. Each $\beta_{pk}$ represents the average partial regression coefficient relating the predictor of interest to students' rate of change, and $u_{pi}$ is the residual between the fitted predictor value for each student's rate of change and the observed rate of change.

Our final analyses provided empirical benchmarks (Bloom et al., 2008) of achievement scores as a contextual aid in interpreting students' mathematics growth trajectories. We used two methods to provide interpretive benchmarks: (a) year-to-year growth ESs for each student group, and (b) achievement gap ESs between student exceptionality groups and all students without disabilities (i.e., GE and AIG students combined). We estimated year-to-year ESs for GE students, AIG students, and students from each exceptionality category by examining the mean difference from one year to the next in ratio to the pooled standard deviation for the two years. To estimate achievement gap ES, the mean difference between the exceptionality group and SWoD students was computed and divided by the standard deviation of the scores for all students in that grade (Bloom et al., 2008). Both ESs were Cohen's *d* statistic, and were interpreted using the same rules of thumb regarding ES magnitude as described earlier for Cohen's *h* (Cohen, 1992).

## 3. Results

We used the analytic sample to describe achievement by grade for each group of SWDs, estimate growth trajectories using multilevel methods, and create empirical benchmarks to further characterize mathematics growth and achievement gaps by student group.

### *3.1. Average observed growth by exceptionality category*

Table 3 shows means and standard deviations of mathematics scale scores by grade for all students together and for each student group that had a sample size greater than 100 in 2000–2001. It can be seen in Table 3 and Fig. 1 that most typical trajectories are curvilinear in form, with decelerating growth across grades. It can also be seen in Table 3 that all student groups showed increasing variability in mathematics scores across grades. As expected, the AIG students had the highest mean achievement at initial status in Grade 3 and showed a somewhat higher slope than other student groups. In terms of level of performance, the order of growth trajectories from highest to lowest performing for the remaining student groups were GE, speech-language impairment, hearing impairment, autism, specific learning disability, other health impairment, emotional disturbance, and mild intellectual disability. Although there was some crossing of growth trajectories in the middle of the distribution, generally there was some differentiation

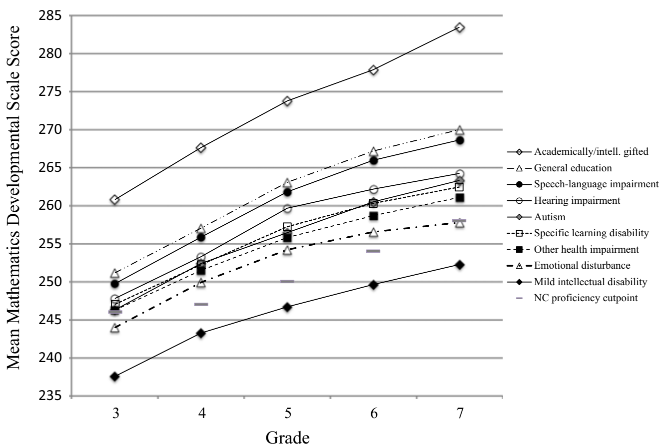


**Fig. 1.** Observed mean mathematics achievement by grade and student group.

between groups of students in initial status. All student groups demonstrated positive growth over time, although there are differences apparent in both rate of change and rate of curvature for some groups. These observations are tested further using the multilevel statistical models described in the next section. Fig. 1 also shows the North Carolina proficiency standard in mathematics for each grade (as indicated by the gray horizontal line at each grade). Because these standards were not linear across grades, the relation of average performance of the student groups varies over grades with fewer student groups exceeding proficiency on average in Grade 7 in comparison to earlier grades.

## 3.2. Multilevel growth models

### 3.2.1. Unconditional and longitudinal level-1 models

The first model applied was a fully unconditional random effects model that only estimated grand means and variance components. We then applied a two-level longitudinal model that estimated quadratic growth trajectories. Introduction of the quadratic model resulted in statistically significant improvement in model fit ($p < .001$; see first columns of Table 4). Across all students, the estimated mean mathematics scale score in Grade 3 was 251.08. The average linear change was significantly different than zero, at 6.92 points ($z = 520.02$, $SE = 0.01$, $p < .001$). The curvature in the growth function was −0.55 scale score points, which was also significantly different than zero ($z = -163.89$, $SE = 0.003$, $p < .001$). This model allowed each growth trajectory parameter (i.e., intercept, linear, and quadratic) to vary randomly across students. A multi-parameter variance component test indicated that this model provided a better fit to the data than a fixed effects model, $\chi^2$ (5) = 46,106.21, $p < .001$, demonstrating that students differed in their growth trajectories. Intercorrelations of the model parameters between intercept and linear, intercept and curvilinear, and linear and curvilinear parameters were .28, .18, and −.61, respectively.

**Table 4**
Fixed and random effects longitudinal HLM regression models, Grades 3 to 7.

| Random effect | Unconditional | | | Specific exceptionality | | | Exceptionality & demographics | | |
|---|---|---|---|---|---|---|---|---|---|
| | Intercept | Linear | Quadratic | Intercept | Linear | Quadratic | Intercept | Linear | Quadratic |
| Mean | 251.08** | 6.92** | −0.55** | 250.95** | 7.02** | −0.56** | 253.14** | 6.99** | −0.54** |
| | (0.03) | (0.01) | (0.00) | (0.03) | (0.01) | (0.01) | (0.04) | (0.02) | (0.01) |
| Academic/intellectually gifted | | | | 9.99** | −0.12* | 0.23** | 7.02** | −0.20** | 0.17** |
| | | | | (0.07) | (0.05) | (0.01) | (0.07) | (0.05) | (0.01) |
| Autism | | | | −6.38** | −0.90* | 0.16 | −8.16** | −0.83 | 0.14 |
| | | | | (0.70) | (0.45) | (0.11) | (0.64) | (0.45) | (0.11) |
| Emotional disturbance | | | | −7.80** | −0.35 | −0.21** | −5.35** | −0.25 | −0.14* |
| | | | | (0.29) | (0.22) | (0.05) | (0.27) | (0.22) | (0.05) |
| Hearing impairment | | | | −3.81** | 0.19 | −0.18 | −4.14** | 0.21 | −0.18 |
| | | | | (0.63) | (0.37) | (0.09) | (0.53) | (0.36) | (0.09) |
| Mild intellectual disability | | | | −14.00** | −1.69** | 0.14** | −10.65** | −1.58** | 0.22** |
| | | | | (0.17) | (0.15) | (0.04) | (0.18) | (0.15) | (0.04) |
| Other health impairment | | | | −5.35** | −1.28** | 0.05 | −5.63** | −1.21** | 0.05 |
| | | | | (0.21) | (0.14) | (0.04) | (0.19) | (0.14) | (0.03) |
| Specific learning disability | | | | −4.48** | −0.86** | 0.01 | −4.16** | −0.79** | 0.03 |
| | | | | (0.10) | (0.06) | (0.02) | (0.09) | (0.06) | (0.02) |
| Speech-language impairment | | | | −1.45** | 0.16 | −0.04 | −1.81** | 0.21* | −0.04** |
| | | | | (0.16) | (0.08) | (0.02) | (0.14) | (0.08) | (0.02) |
| Gender | | | | | | | −0.44** | 0.02 | 0.03** |
| | | | | | | | (0.04) | (0.03) | (0.01) |
| Limited English | | | | | | | −2.69** | −0.13 | 0.04 |
| | | | | | | | (0.16) | (0.11) | (0.03) |
| Parental education | | | | | | | 1.25** | 0.05** | 0.03** |
| | | | | | | | (0.02) | (0.01) | (0.01) |
| Free/reduced lunch | | | | | | | −1.25** | −0.19** | −0.01 |
| | | | | | | | (0.05) | (0.03) | (0.01) |
| Asian | | | | | | | 0.20 | 1.28** | −0.08* |
| | | | | | | | (0.16) | (0.10) | (0.03) |
| Black | | | | | | | −4.13** | 0.22** | −0.11** |
| | | | | | | | (0.05) | (0.03) | (0.01) |
| Hispanic | | | | | | | −0.94** | 0.93** | −0.15** |
| | | | | | | | (0.12) | (0.08) | (0.02) |
| American Indian | | | | | | | −1.85** | −1.46** | 0.29** |
| | | | | | | | (0.17) | (0.12) | (0.03) |
| Variance component | 49.50** | 1.37** | 0.14** | 37.73** | 1.29** | 0.13** | 27.15** | 1.19** | 0.13** |
| Residual | 10.60 | | | 10.59 | | | 10.59 | | |
| Pseudo-$R^2$ (as %) | – | – | – | 23.78 | 5.84 | 7.14 | 45.15 | 13.14 | 7.14 |
| Model *df* | 92,044 | | | 92,036 | | | 92,028 | | |
| ΔDeviance, $\chi^2$ (*df*, *p*-value) | – | | | 24,671.43 (24, <.001) | | | 29,111.25 (24, <.001) | | |

*Note*. Standard errors shown in parentheses.
* $p < .05$
** $p < .001$

### 3.2.2. Specific exceptionality model

We next applied two conditional models that added predictors to the quadratic unconditional model. In the first model, we added dummy coded predictors that reflected students' SWoD group or SWD exceptionality category. Multilevel model results for the exceptionality predictors are shown in the middle columns of Table 4. The average scale score of the GE student reference group (i.e., the group coded zero on all vectors) was about 251. It can be seen that all differences in intercept in Grade 3 between GE students, AIG students, and students in each exceptionality category were statistically significant ($p < .001$). Although AIG students scored about 10 scale score points higher in Grade 3 than GE students, students from all of the exceptionality categories had significantly lower initial mathematics performance. Students classified with a mild intellectual disability showed the largest differences in initial status—on average 14 scale score points lower than GE students. Students with speech-language impairment showed the smallest contrast; they scored an average of 1.45 scale score points lower in Grade 3.

The average initial rate of change for GE students was about 7 scale score points. Students from three exceptionality groups (emotional disturbance, hearing impairment, and speech-language impairment) did not differ significantly in linear rate of change in contrast to the GE students. The remaining comparisons between student groups all showed statistically significant ($p < .001$) lower linear growth. The smallest difference observed was for AIG students, but annual decreases in linear growth rate of from 0.86 to 1.69 scale score points were observed for the remaining exceptionality groups.

Examination of the quadratic term results showed that, on average, GE students were estimated to have a statistically significant deceleration in growth rate of −0.56 scale score points from Grades 3 to 4. Students from three groups (AIG, mild intellectual disability, and emotional disturbance) showed statistically significant differences in rate of curvature. Students classified as AIG and with a mild intellectual disability showed a decrease in the rate of deceleration that resulted in lesser rates of −0.33 and −0.42 scale score points, respectively. In comparison to GE students, students with emotional disturbance showed an additional deceleration in growth rate (−0.21) that resulted in a decrease of −0.77 scale score points. No other group contrasts of the quadratic term were statistically significant.

Variance components, pseudo-$R^2$ (expressed as the percent of variance explained), and deviance statistics are presented at the bottom of Table 4 and show that addition of the exceptionality predictors accounted for approximately 24%, 6%, and 7% of the variance in student intercepts, slopes, and curvature, respectively, in comparison to the unconditional longitudinal model. Comparison of model deviances between the unconditional longitudinal and specific exceptionality models resulted in a statistically significant reduction in unexplained variance, $\chi^2$ (24) = 24,671.43, $p < .001$. Intercorrelations of the model parameters between intercept and linear, intercept and curvilinear, and linear and curvilinear parameters were .26, .14, and −.63, respectively.

### 3.2.3 . Exceptionality and demographics model

We then expanded the multilevel growth model by adding an additional set of predictors representing student demographics and background characteristics (see right-most columns of Table 4). The estimated mean initial status or intercept (253.14) now represented the average mathematics achievement in Grade 3 for students who were White males, GE, English proficient, not receiving free/reduced lunch, and had parents at the midpoint of education level. Although the magnitude of several exceptionality parameter estimates changed through addition of the socio-demographic predictors, hypothesis testing results for the mean of parameters and for specific exceptionality category contrasts were very similar to those for the previous model with three exceptions: linear change for the autism group was no longer statistically significant, and both linear change and curvature for students with speech-language impairments were now statistically significant in this model when they had not been significant in the previous model. Estimated growth trajectories for this final model are still well represented by those shown in Fig. 1 with the same trends and ordering of student groups; however, modeled trajectories were smoother and more linear in appearance.[5]

Examination of results for the added predictors showed that, controlling for all other predictors, females (−0.44), limited English proficient students (−2.69), free/reduced lunch recipients (−1.25), Black students (−4.13), Hispanic students (−0.94), and American Indian students (−1.85) had significantly lower initial mathematics performance than the reference group in Grade 3. Results for the linear growth parameter showed that all predictors except student gender and limited English proficiency showed statistically significant differences in rate of change in comparison to the reference group. Controlling for all other predictors, students with higher levels of parental education (0.05), who were Asian (1.28), Black (0.22), or Hispanic (0.93) showed larger rates of change than the reference group; and students who were free/reduced lunch recipients (−0.19) or American Indians (−1.46) showed lesser rates of change. Results for the quadratic parameter showed that the limited English proficient students and free/reduced lunch recipients did not differ significantly from the reference group. Controlling for all other predictors, female students (0.03), those with higher levels of parental education (0.03), and American Indian students (0.29), showed acceleration in growth rate, whereas free/reduced lunch recipients (−0.01), Asian (−0.08), Black (−0.11), and Hispanic (−0.15) students all showed statistically significant deceleration in growth rate compared to the reference group.

Examination of pseudo-$R^2$ (expressed as the percent of variance explained) and deviance statistics showed that addition of the socio-demographic predictors accounted for approximately 45%, 13%, and 7% of the variance in student intercepts, slopes, and curvature, respectively in comparison to the unconditional longitudinal model. Examination of pseudo-$R^2$ between Model 2 and Model 3 showed that addition of the socio-demographic predictors accounted for an additional 28% and 8% of variance in intercepts and slopes respectively; there was no change in explained variance of curvature parameters. Comparison of model deviances between the unconditional longitudinal and specific exceptionality and demographics models resulted in a statistically significant reduction in

[5] Additional figures and tables based on HLM estimated model results are available on request from the first author.

unexplained variance, $\chi^2$ (48) = 53,782.68, $p < .001$, and comparison of model deviances between Model 2 (specific exceptionality) and Model 3 (specific exceptionality and demographics) also resulted in a statistically significant reduction in unexplained variance, $\chi^2$ (24) = 29,111.25, $p < .001$.

Intercorrelations of the model parameters between intercept and linear, intercept and curvilinear, and linear and curvilinear parameters were .29, .04, and −.65, respectively. Thus, after controlling for demographic characteristics, there was a modest correlation between intercept and linear slope indicating a Matthew effect relation in which, on average, those who scored higher in Grade 3 grew at a more rapid rate and those with lower Grade 3 scores grew at a slower rate. There was no relation of intercept to curvature. And there was a strong negative relation between linear and curvilinear growth parameters in which those with higher initial rates of growth had the greatest deceleration in growth and those with lower initial growth had less moderation of growth rate over time.

One of our research questions involved evaluating the impact of inclusion of socio-demographic predictors. Estimates from the middle to the right column of Table 4 were compared to determine the effect of controlling for socio-demographic characteristics on estimates of exceptionality group differences. The magnitude of the change was evaluated by comparing the size of coefficient change to the coefficient's standard error. It can be seen that, after controlling for socio-demographic characteristics, initial level of mathematics performance (i.e., intercept) increased substantially (>2 *SE*) for students with mild intellectual disability, emotional disturbance, and specific learning disabilities. Initial level of mathematics performance decreased substantially (>2 *SE*) for AIG students, autism, and speech-language impairment. None of the differences in initial growth rate estimates (i.e., linear) were larger than two standard errors after controlling for socio-demographic characteristics. Examination of changes in acceleration and deceleration of growth (i.e., quadratic) showed that AIG students showed less acceleration (>2 *SE*) in mathematics growth and students with mild intellectual disability showed greater acceleration (>2 *SE*) in mathematics growth over time after controlling for socio-demographic characteristics.

### *3.3. Empirical growth benchmarks*

To provide additional empirical context for interpretation of differences in student growth we examined two representations of the ES of student group differences: (a) growth in mathematics performance from year-to-year and (b) achievement gaps between student exceptionality groups and all SWoDs.

#### *3.3.1. Mathematics growth effect sizes*

Table 5 and Fig. 2 show year-to-year growth expressed as an ES for all SWoDs, GE students, AIG students, and students in each exceptionality category. As reported in the first column of Table 5, student growth from Grade 3 to Grade 4 is characterized by large ESs for all student groups, ranging from 0.72 for the other health impairment group to 1.19 for the mild intellectual disability group. Inspection of the grade transition ESs showed that they generally diminished across grades. For all SWoDs, this trend results in a decrease in ES from 0.78 in grade transition 3–4 to 0.32 in grade transition 6–7. This decrease in year-to-year growth is generally reflected for all student groups and is apparent in Fig. 2. AIG students and students with mild intellectual disability maintained relatively higher year-to-year growth than other student groups. However, it is also worth noting that students with mild intellectual disability showed the largest deceleration in growth over grades.

#### *3.3.2. Achievement gap effect sizes*

Another representation of mathematics achievement for student groups that is particularly policy relevant in the context of NCLB can be obtained by inspecting the relative growth of one student group in comparison to others. The usual comparison made in NAEP reporting and in the context of NCLB accountability reports is between SWD and SWoD student groups. However, as we have described, there is merit in further disaggregation of SWD into specific exceptionality groups as well as the disaggregation of SWoDs into GE and AIG students. Thus, we report achievement gap results using the common approach with SWoDs as the reference group (GE and AIG together), but we also report achievement gaps with SWoDs disaggregated using only GE students as the reference

**Table 5**
Mathematics growth effect size over time by student group.

| | Grade transition | | | |
|---|---|---|---|---|
| Student group | 3–4 | 4–5 | 5–6 | 6–7 |
| Students without disabilities | 0.78 | 0.74 | 0.48 | 0.32 |
| General education | 0.81 | 0.78 | 0.50 | 0.31 |
| Academically/intell. gifted | 1.15 | 0.99 | 0.65 | 0.72 |
| Students with disabilities | 0.74 | 0.54 | 0.32 | 0.24 |
| Autism | 0.75 | 0.42 | 0.39 | 0.25 |
| Emotional disturbance | 0.84 | 0.54 | 0.27 | 0.14 |
| Hearing impairment | 0.75 | 0.81 | 0.29 | 0.22 |
| Mild intellectual disability | 1.19 | 0.67 | 0.55 | 0.47 |
| Other health impairment | 0.72 | 0.54 | 0.34 | 0.26 |
| Specific learning disability | 0.75 | 0.66 | 0.37 | 0.25 |
| Speech-language impairment | 0.78 | 0.70 | 0.46 | 0.27 |

*Note*. Students without disabilities group is composed of the combination of general education and academically/intellectually gifted students.

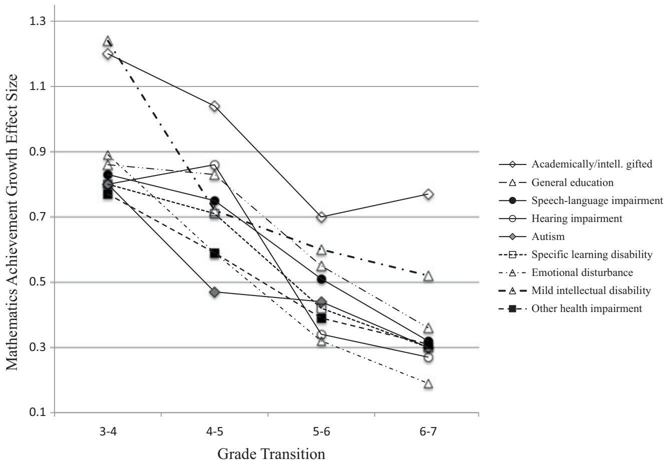


**Fig. 2.** Mathematics achievement growth effect size at each grade transition by student group.

group. Table 6 documents ES comparisons of students by exceptionality group with all SWoDs together in the upper portion of the table and in comparison only with GE students in the lower portion of the table. In comparison to SWoDs, all student exceptionality group mathematics scores were lower by 0.30 to almost 2.0 standard deviations in Grade 3. These results using aggregated SWoD students are also shown in Fig. 3. Although there is variation from grade to grade with some narrowing for several groups in Grades 4 and 7 and some widening for several groups in Grades 5 and 6, achievement gap differences were relatively stable from Grade 3 to Grade 7 and remain from 0.23 to 1.74 standard deviations lower than SWoD students in Grade 7. In the lower portion of the table, it can be seen that, in comparison to GE students only, the same patterns are apparent although the magnitude of achievement gaps is smaller. It is also apparent that AIG students have a sizeable positive achievement gap over GE students of more than one standard deviation across all grades.

## 4. Discussion

Although there is substantial interest in mathematics achievement growth and gaps for different student groups, much of our research-based knowledge has been from small, nonrepresentative samples (e.g., Geary et al., 2012), national databases (Carlson et al., 2011; Morgan et al., 2009; Wei et al., 2013), and aggregated groups. Although this research provides useful information, the achievement growth of student groups on state assessments, especially students receiving special education services, has largely been unexplored. The present study addressed these limitations in the existing literature by examining children's mathematics growth trajectories over multiple grades and is the first study that reports growth results for multiple, specific student exceptionality groups on an operational, state-wide achievement test used for accountability reporting. Our results estimated growth trajectories for GE students, AIG students, and students in seven disability categories over Grades 3 to 7 as well as evaluating the relations of student

**Table 6**
Mathematics achievement gap effect sizes by exceptionality category in comparison to all students without disabilities and to general education students.

| | Grade | | | | |
|---|---|---|---|---|---|
| Student group | 3 | 4 | 5 | 6 | 7 |
| vs. all students without disabilities | | | | | |
| Autism | −0.76 | −0.68 | −0.87 | −0.83 | −0.72 |
| Emotional disturbance | −1.07 | −0.98 | −1.13 | −1.26 | −1.23 |
| Hearing impairment | −0.56 | −0.57 | −0.50 | −0.64 | −0.64 |
| Mild intellectual disability | −1.93 | −1.79 | −1.99 | −2.01 | −1.74 |
| Other health impairment | −0.76 | −0.79 | −0.94 | −1.02 | −0.92 |
| Specific learning disability | −0.66 | −0.70 | −0.78 | −.085 | −0.80 |
| Speech-language impairment | −0.30 | −0.26 | −0.25 | −0.23 | −0.23 |
| vs. general education students | | | | | |
| Academically/intell. gifted | +1.28 | +1.29 | +1.23 | +1.16 | +1.24 |
| Autism | −0.65 | −0.57 | −0.76 | −0.73 | −0.61 |
| Emotional disturbance | −0.96 | −0.87 | −1.02 | −1.16 | −1.12 |
| Hearing impairment | −0.45 | −0.46 | −0.39 | −0.54 | −0.53 |
| Mild intellectual disability | −1.82 | −1.68 | −1.89 | −1.91 | −1.63 |
| Other health impairment | −0.65 | −0.68 | −0.84 | −0.93 | −0.82 |
| Specific learning disability | −0.55 | −0.59 | −0.67 | −0.75 | −0.69 |
| Speech-language impairment | −0.19 | −0.14 | −0.15 | −0.13 | −0.13 |

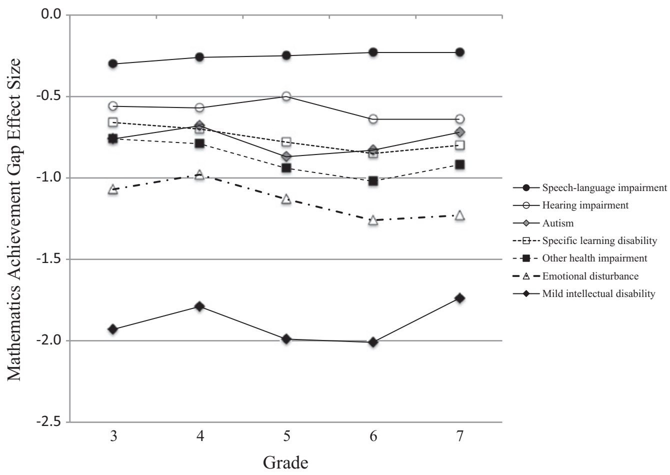


**Fig. 3.** Achievement gap effect sizes between all SWoD and each exceptionality group by grade.

socio-demographic characteristics to growth. Although many studies have investigated achievement gap differences, few provide empirical benchmarks of these differences by reporting ESs. In the present study, two recommended ES benchmarks of change (Bloom et al., 2008) were reported: year-to-year growth ESs and achievement gap ESs that reflect the goals of NCLB accountability policy.

### *4.1. Major study findings in relation to previous research*

On the North Carolina EOG Mathematics Tests, we found that achievement growth over Grades 3 to 7 was best represented as a curvilinear function with achievement growth decelerating over time. However, the rate of curvature was quite small both in an absolute sense for scale score units and in terms of the percentage of variance explained by the curvature term. These results generally confirm those of other recent studies (Bloom et al., 2008; Morgan et al., 2009; Morgan et al., 2011; Shin et al., 2013; Wei et al., 2013) that all reported curvilinear mathematics growth functions with decelerating growth over age or grade. Not surprisingly given different samples and instruments, there are some differences across studies in rates of growth or the timing of the deceleration of growth. For example, Wei et al. (2013) found that growth slowed at approximately age 13 across all disability groups whereas deceleration was apparent in our results approximately two years earlier for most student groups. The magnitude of changes in growth described in other published research is less apparent and more difficult to compare because those studies did not report ES statistics.

To address our first two research questions, we examined growth for specific student groups. Overall, the same pattern of statistically significant but diminishing growth across grades was observed for almost all groups. Two exceptions were students with mild intellectual disabilities who showed larger growth ESs over grades than other groups and AIG students whose growth ES from Grades 6 to 7 actually increased rather than decreased. However, there was also statistically significant variability by group—both in terms of initial mathematics achievement and in rates of growth over time. In terms of initial achievement level, all groups of students with disabilities showed lower achievement than their peers without disabilities. Among the special education groups, students with speech-language impairments were the highest performing, and students with mild intellectual disability were the lowest performing.

These findings are similar to previous research (Morgan et al., 2011; Shin et al., 2013; Wei et al., 2013) in the general ordering of group performance but more detailed than previous research in showing the performance of multiple exceptionalities as well as AIG student performance in contrast to GE students. The extant findings regarding differences in mathematics growth across grade by exceptionality group have been more variable. For example, Wei et al. (2013) reported almost no differences in mathematics growth over time among different exceptionality groups in comparison to students with learning disabilities; Morgan et al. (2011) found that mathematics growth for students with speech-language impairments did not differ from growth for SWoDs, but the growth for students with learning disabilities was slower. Our study extended these findings in three important ways. First, we showed the pattern of growth trajectories for student groups on an operational statewide accountability test. Second, we added to the small literature that has examined mathematics growth for AIG students. Last, our results provided additional context for comparison and understanding of growth for exceptional children by reporting ESs and through the provision of two specific SWoD reference groups, GE and AIG students.

In the limited literature on the academic growth of AIG students, there has been a concern that NCLB proficiency-based assessments are unlikely to adequately measure their growth (McCoach, Rambo, & Welsh, 2013) because of limitations in content coverage, type of tests, and test ceilings that are too low to measure progress or growth of AIG students (Subotnik et al., 2011). In the present study, mathematics growth of a large sample of AIG students was modeled, and in comparison to GE students, we found significantly higher initial performance in Grade 3, significantly slower linear growth, but significantly less deceleration in growth rate. Although this study had no means to evaluate the presence or impact of ceiling (or floor) effects on student groups, it is noteworthy that there is positive growth for all groups over grades despite any measurement artifacts that may have been present and AIG growth actually accelerates from Grade 6 to Grade 7.

To address our third research question, we examined the relation between socio-demographic characteristics of students and mathematics growth. We found statistically significant differences in mathematics performance as a function of student demographic characteristics with higher performance associated with male, White, higher parental education, no free/reduced lunch subsidy, and native English speakers. These results are consistent with a large body of research that establishes relations of these student characteristics with student achievement (Holman, 1995; Lee & Burkam, 2002; Phillips, Norris, Osmond, & Maynard, 2002; Tate, 1997).

Studies investigating the growth trajectories of SWDs have not always statistically controlled for variation that might be related to student demographic characteristics; however, there is evidence that including such controls significantly alters exceptionality group estimates (e.g., Kieffer, 2011; Morgan et al., 2011; Wei et al., 2011, 2013). In the present study, we found that addition of student socio-demographic characteristics resulted in substantial changes in the magnitude of intercept coefficients for most exceptionality categories with some increases and some decreases in the difference between the achievement of exceptional and GE students. After controlling for socio-demographic characteristics, we found no substantial changes in the size of exceptionality coefficients for linear growth and only two differences for rate of curvature: less acceleration for AIG students and greater acceleration for students with mild intellectual disability. These results suggest that exceptionality group differences, especially initial status, are confounded to some extent by differential representation of exceptional students in particular socio-demographic groups. This confounding has been reported in previous research (Blackorby et al., 2005; Coutinho, Oswald, & Best, 2002; Losen & Orfield, 2002), and as shown in Table 2 for this sample, there was disproportional representation with the SWD group composed of larger percentages of males, free-reduced lunch students, and students whose parents had lower levels of education.

Finally, we addressed our research questions about achievement gaps by calculating estimates of year-to-year growth ES for each student group and by examining achievement gaps between exceptionality groups and all SWoDs. By using these methods, we provided results in a metric that is comparable across studies and independent of the particular statistical model applied. There is evidence that there are achievement gaps when students enter school and at the time of their first assessment on state mandated accountability tests (Lee & Burkam, 2002). Our estimates of growth ES showed that, for all student groups, the magnitude of growth from earlier grades into middle school decreased, which further underscores the curvilinear nature of mathematics growth described earlier.

Results conflict on the stability of the mathematics achievement gap over time. Although some researchers have reported widening of the achievement gap for special education students in aggregate or in selected exceptionality groups (e.g., Jordan et al., 2002; Morgan et al., 2011), our results are more consistent with studies that reported stable achievement gaps over time (e.g., Shin et al., 2013). Less common are studies that report achievement gap differences for multiple, specific student exceptionality categories, and this study is the first to have reported achievement gap differences with AIG students differentiated from GE students. By examining AIG students separately as in Table 6, a more nuanced picture of the gap between specific exceptionality groups and GE students emerged. Achievement gaps between SWD and GE students are, of course, smaller when AIG students are considered separately. Although it is important to examine the individual performance of AIG students, it may also be that differences found among previous studies of academic achievement gaps are complicated by differences in the composition or performance of AIG students in the combined SWoD group.

It should also be noted that many studies that describe achievement gaps and almost all studies of achievement gaps for specific exceptionality groups have depended on descriptive statistics and visual inspection of results to characterize the magnitude of achievement gaps. We believe that conclusions drawn from these studies might change if the appropriate standard deviations were taken into account in describing mean differences using an ES. Use and reporting of ES descriptions of achievement gap differences can facilitate comparisons of results across studies and yield richer interpretation of the magnitude of differences (Bloom et al., 2008). Our results showed that, in comparison to SWoDs, achievement gaps ranged from relatively small differences for students with speech-language impairments to large differences for students with mild intellectual disability, and although there was some variation across grades, there was no consistent evidence that achievement gaps were closing over Grades 3 to 7.

### 4.2. Limitations

A number of limitations in the present study should be noted and considered when interpreting results. First, our analyses did not include a variety of additional student, family, or school characteristics that may be related to student academic achievement (e.g., resources in the home, income level, family stability, intellectual ability, student engagement, and school climate). Our analyses also only examined the partial regression effects of socio-demographic predictors rather than complete exploration of interaction effects, which were beyond the scope of the present study. Nonetheless, by including socio-demographic predictors in our analytic model, we controlled for student characteristics to some extent. However, the issue of disproportional representation of students with certain characteristics is important and may require more careful and extensive study to fully understand the ways in which student background and characteristics interact with achievement growth. We also were unable to adequately represent and model several IDEIA exceptionality categories (i.e., autism, orthopedic impairment, traumatic brain injury, and visual impairment) because of small sample size, and our results are limited by a number of case exclusions for missing values. As in all longitudinal studies, some attrition occurred over time, although the rate in the present study (approximately 4% per year) was modest in comparison to rates reported in some other longitudinal studies (e.g., McCaffrey, Lockwood, Mariano, & Setodji, 2005; Zvoch & Stevens, 2005).

Our sample also does not reflect the substantial temporal complexity that exists in characterizing students as having "special education" status or as belonging to a specific exceptionality group. Like most published longitudinal studies of academic achievement, for the purpose of analysis, student characteristics were defined at wave 1 (Grade 3). However, we know that students enter and exit into special education over time, may change exceptionality classification, may sometimes take the regular state assessment and sometimes an alternate assessment, and are more likely to be retained in grade than GE students (Schulte, 2010; Schulte & Stevens, in press; Ysseldyke & Bielinski, 2002). Analyses that do not model these complex changes in student status and classification over time may result in biased estimates. It is also important to note that, although we contributed substantial new information through the use of one state's accountability test and the reporting of ES benchmarks, the results reported here may not generalize to other state testing and educational systems. However, in contrast, it is important to note as well that most research on mathematics achievement growth and gaps has used national surveys that are unlikely to generalize in all respects to state achievement tests or to represent the sample composition of students engaged in high-stakes educational accountability testing. Both kinds of research are valuable, one nationally representative, the other representative of a state testing system, but it should not be assumed that generalizations can be made freely from one to another without an empirical basis.

### *4.3. Conclusions and future directions*

Our general findings are largely consistent with and complementary to many previously published findings showing the significant but decelerating mathematics achievement growth of students over grades. The results of this study, however, also fill in a number of important gaps in the literature by analyzing an operational state test, providing well defined comparison groups, and expressing differences in growth over grades and differences between groups using ES benchmarks. Our findings also extend a very small literature that examines achievement growth for multiple student exceptionality groups, including AIG students. Additional research is needed that expands on these results with the same and additional exceptionality groups. It is also important that some portion of research on student achievement growth trajectories is conducted with state assessment data used for NCLB accountability reporting so that educational program planning, decision-making, and policy formation can be informed and enhanced. It is unlikely that our results will perfectly replicate to other state assessment instruments or states whose student population differs in the composition of racial/ethnic groups, percentage of English language learners, and rates of economic disadvantage. It is therefore important, given the high stakes surrounding accountability testing, that we more fully understand the growth of all students on state tests as well as the growth of specific groups of students including exceptional children. For practitioners, our results add important information by showing the performance of specific groups of children.

Wei et al. (2013) describe the NCLB approach to classification of special education students as a "one size fits all" approach. Our results confirmed that perspective and showed substantial variation in growth trajectories by student group that may be masked by the aggregation of groups into a single special education category. Although many similarities were found in growth trajectories for students in different exceptionality groups, average achievement gaps ranged from approximately one-quarter standard deviation for students with speech-language impairments to about two standard deviations for students with mild intellectual disability. Perhaps less obvious than this heterogeneity in student group performance are the implications for practitioners and policy makers. Student heterogeneity requires greater attention to instructionally sensitive student assessment and different approaches to instructional intervention to foster academic improvement and ensure that all students learn to proficiency (Kauffman, Bantz, & McCullough, 2002; Tomlinson et al., 2003).

The laudable goal of closing the gap between SWDs and SWoDs is an important policy target. However, the policy may not fully embrace empirical evidence about student growth. For those students who are "behind" to catch up with other students, it is necessary that they not only make academic progress but that they make progress at faster growth rates than their higher performing and often more advantaged peers. Results presented here show that, in fact, SWDs generally are not closing the achievement gap relative to SWoDs. As noted by others (e.g., Morgan et al., 2011; Vaughn & Wanzek, 2014; Wei et al., 2013), these and similar findings raise important questions about how much growth should be expected for these students, what are reasonable expectations for "proficiency" and for narrowing the achievement gap, and to what extent the gap between current and ideal educational practices versus persistent cognitive and behavioral impairments for SWDs contribute to the observed outcomes.

Expectations for growth cannot rest solely on policy intent or a desire to set lofty goals but also must be contextualized by empirical findings that describe the growth that occurs for students as a function of their exceptionality and background. We believe, through the use of ES interpretations of student growth, the results presented here provide rich additional context for characterizing and interpreting student growth for SWDs and SWoDs, including both GE and AIG students. It is important to express not only how much growth we would like to see from a pedagogical or policy perspective but also how much growth typically occurs; how that growth is conditioned on student exceptionality, characteristics, and background; and what are reasonable expectations for future student achievement.

## References

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology*, *34*, 1373–1399.

Blackorby, J., Wagner, M., Cameto, R., Levine, P., Davies, E., et al. (2005). *Engagement, academic, social adjustment, and independence: The achievements of elementary and middle school students with disabilities.* Menlo Park, CA: SRI International.

Blair, C., & Scott, K. G. (2002). Proportion of LD placements associated with low socioeconomic status: Evidence for a gradient? *Journal of Special Education*, *36*, 14–22.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*, 289–328.

Carlberg, C., & Kavale, K. (1980). The efficacy of special versus regular class placement for exceptional children: A meta-analysis. *Journal of Special Education*, *14*, 295–309.

Carlson, E., Jenkins, F., Bitterman, A., & Keller, B. (2011). *A longitudinal view of receptive vocabulary and math achievement of young children with disabilities, (NCSER 2011-3006).* U.S. Department of Education. Washington, DC: National Center for SpecialEducation Research.
Center on Education Policy (2009). *State test score trends through 2007–08, part 4: Has progress been made in raising achievement for students with disabilities?* Washington, DC: Author.
Chatterji, M. (2005). Achievement gaps and correlates of early mathematics achievement: Evidence from the ECLS K–First Grade sample. *Education Policy Analysis Archives*, *13*(46).
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.
Council for Exceptional Children (2013). NAEP results show wide achievement gaps between students with, without disabilities. Retrieved from. http://www.policyinsider.org/2013/11/naep-results-show-wide-achievement-gaps-between-students-with-without-disabilities.html (November 14)
Council for Exceptional Children—The Association for the Gifted (n.d.). Growth in achievement of advanced students. Retrieved from. http://cectag.com/wp-content/uploads/2012/04/Growth-Models.pdf.
Coutinho, M. J., Oswald, D. P., & Best, A. M. (2002). The influence of sociodemographics and gender on the disproportionate identification of minority students as having learning disabilities. *Remedial and Special Education*, *23*, 49–59.
Davison, M. L., Seok Seo, Y., Davenport, E. C., Jr., Butterbaugh, D., & Davison, L. J. (2004). When do children fall behind? What can be done? *Phi Delta Kappan*, *85*, 752–761.
Denton, K., & West, J. (2002). *Children's reading and mathematics achievement in kindergarten and first grade.* Washington, DC: National Center for Education Statistics.
Ding, C. S., Davison, M. L., & Petersen, A. C. (2005). Multidimensional scaling analysis of growth and change. *Journal of Educational Measurement*, *42*, 171–191.
Eckes, S. E., & Swando, J. (2009). Special education subgroups under NCLB: Issues to consider. *Teachers College Record*, *111*, 2479–2504.
Enders, C. K. (2010). *Applied missing data analysis.* New York, NY: Guilford Press.
Galindo, C. (2010). English language learners' mathematics and reading achievement trajectories in the elementary grades. In E. García, & E. Frede (Eds.), *Developing the research agenda for young English language learners* (pp. 42–58). New York, NY: Teachers College Press.
Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology*, *104*, 206–223.
Han, W. J. (2008). The academic trajectories of children of immigrants and their school environments. *Developmental Psychology*, *44*, 1572–1590.
Hemphill, F. C., Vanneman, A., & Rahman, T. (2011). *Achievement gaps: How Hispanic and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress (NCES 2011-459).* Washington, DC: National Center for Education Statistics.
Holman, L. J. (1995). Impact of ethnicity, class, and gender on achievement of border area students on a high-stakes examination. *Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA* (April).
Hyde, J. S., Fennema, E., & Lamon, S. J. (1998). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139–155.
Individuals with Disabilities Education Improvement Act, H.R. 1350, 108th Congress (2004).
Jordan, N. C., Kaplan, D., & Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two year longitudinal study. *Journal of Educational Psychology*, *94*, 586–597.
Judge, S., & Watson, S. M. R. (2011). Longitudinal outcomes for mathematics achievement for students with learning disabilities. *Journal of Educational Research*, *104*, 147–157.
Kauffman, J. M., Bantz, J., & McCullough, J. (2002). Separate and better: A special public school class for students with emotional and behavioral disorders. *Exceptionality*, *10*, 149–170.
Kieffer, M. J. (2011). Converging trajectories: Reading growth in language minority learners and their classmates, kindergarten to grade 8. *American Educational Research Journal*, *48*, 1187–1225.
Lee, J. (2010). Tripartite growth trajectories of reading and mathematics achievement: Tracking national academic progress at primary, middle and high school levels. *American Educational Research Journal*, *47*, 800–832.
Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate.* Washington, DC: Economic Policy Institute.
LoGerfo, L., Nichols, A., & Reardon, S. (2006). *Achievement gains in elementary and high school.* Washington, DC: Urban Institute.
Lohman, D. F., & Korb, K. A. (2006). Gifted today but not tomorrow? Longitudinal changes in ability and achievement during elementary school. *Journal for the Education of the Gifted*, *29*, 451–484.
Losen, D., & Orfield, G. (2002). *Racial inequity in special education.* Cambridge, MA: Harvard Education Publishing Group.
Lubinski, D., Webb, R. M., Morelock, M. J., & Benbow, C. P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology*, *86*, 718–729.
Ma, X. (2005). A longitudinal assessment of early acceleration of students in mathematics on growth in mathematics achievement. *Developmental Review*, *25*, 104–132.
Manna, P., & Ryan, L. L. (2011). Competitive grants and educational federalism: President Obama's Race to the Top program in theory and practice. *Publius: The Journal of Federalism*, *41*, 522–546.
McCaffrey, D. F., Lockwood, J. R., Mariano, L. T., & Setodji, C. (2005). Challenges for value-added assessment of teacher effects. In R. Lissitz (Ed.), *Value added models ineducation: Theory and applications* (pp. 111–144). Maple Grove, MN: JAM Press.
McCoach, B. D., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, *98*, 14–28.
McCoach, D. B., Rambo, K. E., & Welsh, M. (2013). Assessing the growth of gifted students. *Gifted Child Quarterly*, *57*, 56–67.
McDonnell, L. M., McLaughlin, M. J., & Morison, P. (Eds.). (1997). *Educating one and all: Students with disabilities and standards-based reform.* Washington, DC: National Academy Press.
Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, *42*, 306–321.
Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities*, *44*, 472–488.
National Early Literacy Panel (2008). *Developing early literacy: Report of the National Early Literacy Panel.* Washington, DC: National Institute of Literacy.
No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107–110, § 115, Stat. 1425 (2002).
North Carolina Department of Public Instruction (1996). *End-of-Grade Tests: Technical report.* Raleigh, NC: Author (Retrieved from http://www.ncpublicschools.org/docs/accountability/NCEndofGradeTests.pdf).
North Carolina Department of Public Instruction (2006). *The North Carolina mathematics tests: Technical report.* Raleigh, NC: Author (Retrieved from http://www.ncpublicschools.org/docs/accountability/testing/mathtechmanual.pdf).
Northwest Evaluation Association (2011). *Measures of academic progress and measures of academic progress for primary grades technical manual.* Portland, OR: Author.
Northwest Evaluation Association (2014). *A study of the alignment of the NWEA RIT scale with the North Carolina State End of Grade (EOG) testing program.* Portland, OR: Author.
Phillips, M., Brooks-Gunn, J., Duncan, G. J., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the black–white test score gap. In C. Jencks, & M. Phillips (Eds.), *The black–white test score gap* (pp. 103–145). Washington, DC: Brookings Press.
Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology*, *94*, 3–13.
Princiotta, D., Flanagan, K. D., & Germino-Hausken, E. (2006). *Fifth grade: Findings from the fifth grade follow-up of the Early Childhood Longitudinal Study, kindergarten class of 1998–99 (ECLS-K).* Washington, DC: U.S. Department of Education.
Protopapas, A., Sideridis, G. D., Mouzaki, A., & Simos, P. G. (2011). Matthew effects in reading comprehension: Myth or reality? *Journal of Learning Disabilities*, *44*, 402–420.

Rambo-Hernandez, K. E., & McCoach, D. B. (2014). High-achieving and average students' reading growth: Contrasting school and summer trajectories. *The Journal of Educational Research*. http://dx.doi.org/10.1080/00220671.2013.850398.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Reardon, S. F., & Galindo, C. (2009). The Hispanic–White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*, 853–891.

Renaissance Learning (2008). *STAR Math computer-adaptive math test: Technical manual*. Wisconsin Rapids, WI: Author.

Sabornie, E. J., Cullinan, D., Osborne, S. S., & Brock, L. B. (2005). Intellectual, academic, and behavioral functioning of students with high-incidence disabilities: A cross-categorical meta-analysis. *Exceptional Children*, *72*, 47–63.

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*, *48*, 115–136.

Scarborough, H. S., & Parker, J. D. (2003). Matthew effects in children with learning disabilities: Development of reading, IQ, and psychosocial problems from grade 2 to grade 8. *Annals of Dyslexia*, *53*, 47–71.

Schulte, A. C. (2010). Assessing growth for students with disabilities on large scale assessments. *Paper presented at the National Conference on Student Assessment, Detroit, MI* (June).

Schulte, A. C., & Stevens, J. J. (2014). Once, sometimes, or always in special education: Do inclusion criteria change portrayals of mathematics growth and achievement gaps for students with disabilities? *Exceptional Children* (in press).

Schulte, A. C., Villwock, D. N., Whichard, S. W., & Stallings, C. F. (2001). High stakes testing and expected progress standards for students with learning disabilities: A five-year study of one district. *School Psychology Review*, *30*, 487–506.

Shaywitz, B. A., Holford, T. R., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly*, *30*, 387–388.

Shin, T., Davison, M. L., Long, J. D., Chan, C. -K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences*, *23*, 92–100.

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*, 3–54.

Tate, W. F. (1997). Race–ethnicity, SES, gender, and language proficiency trends in mathematics achievement: An update. *Journal for Research in Mathematics Education*, *28*, 652–679.

Thissen, D. (2007). Linking assessments based on aggregate reporting: Background and issues. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 287–312). New York, NY: Springer.

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen, & H. Wainer (Eds.), *Test scoring* (pp. 73–140). Mahwah, NJ: Erlbaum.

Thissen, D., Pommerich, M., Billeaud, K., & Williams, V. S. L. (1995). Item response theory for scores on tests including polytomous items with order responses. *Applied Psychological Measurement*, *19*, 39–49.

Thurlow, M., Altman, J., Cormier, D., & Moen, R. (2008). *Annual performance report: 2005–2006 state assessment data*. Minneapolis, MN: University of Minnesota, NCEO.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., et al. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal of the Education of the Gifted*, *27*, 119–145.

U.S. Census Bureau (2014). *Profile of general demographic characteristics: 2000. U.S. Census FactFinder from Census 2000 Summary File 1 (SF 1).* (April, Retrieved from http://factfinder2.census.gov).

U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2013). *National Assessment of Educational Progress (NAEP) data explorer. Custom cross-tabulation of data.* (Retrieved from http://nces.ed.gov/nationsreportcard/naepdata/).

U.S. Department of Education (2009). *Race to the Top executive summary*. Washington, D.C.: Author (Retrieved from the web: http://www2.ed.gov/programs/racetothetop/executive-summary.pdf).

U.S. Department of Education, Digest of Education Statistics (2002). *Table 55. Number and percent of children served under Individuals with Disabilities Act and Chapter 1*. Washington, DC: Author (Retrieved from http://nces.ed.gov/programs/digest/).

Vanneman, A., Hamilton, L., Anderson, J. B., & Rahman, T. (2009). *Achievement gaps: How Black and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress (NCES 2009-455)*. Washington, DC: National Center for Education Statistics.

Vaughn, S., & Wanzek, J. (2014). Intensive intervention in reading for students with reading disabilities: Meaningful impacts. *Learning Disabilities Research & Practice*, *29*, 46–53.

Watson, S. M. R., & Gable, R. A. (2013). Unraveling the complex nature of mathematics learning disability: Implications for research and practice. *Learning Disability Quarterly*, *36*, 178–187.

Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in reading achievement in a national sample of students with disabilities ages 7 to 17. *Exceptional Children*, *78*, 89–106.

Wei, X., Lenz, K. B., & Blackorby, J. (2013). Math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, *34*, 154–165.

Williams, V. S. L., Pommerich, M., & Thissen, D. (1998). A comparison of developmental scales based on Thurstone methods and item response theory. *Journal of Educational Measurement*, *35*, 93–107.

Williams, V. S. L., Rosa, K. R., McLeod, L. D., Thissen, D., & Sanford, E. E. (1998). Projecting to the NAEP scale: Results from the North Carolina End-of-Grade testing program. *Journal of Educational Measurement*, *35*, 277–296.

Wright, W. E., & Li, X. (2008). High-stakes mathematics tests: How No Child Left Behind leaves newcomer English language learners behind. *Language Policy*, *7*, 201–216.

Ysseldyke, J., & Bielinski, J. (2002). Effect of different methods of reporting and reclassification on trends in test scores for students with disabilities. *Exceptional Children*, *68*, 189–200.

Zvoch, K., & Stevens, J. (2005). Sample exclusion and student attrition effects in the longitudinal study of middle school mathematics performance. *Educational Assessment*, *10*, 105–123.