ABSTRACT

EWING, KATHERINE ANNE. Estimating the Effectiveness of Special Education Using Large-Scale Assessment Data. (Under the direction of Ann C. Schulte, Ph.D.).

The inclusion of students with disabilities in large scale assessment and accountability programs has provided new opportunities to examine the impact of special education services on student achievement. Hanushek, Kain, and Rivkin (1998, 2002) evaluated the effectiveness of special education programs by examining students' gains on a large-scale assessment of student achievement in Texas as they entered and exited special education programs. They found that special education placement resulted in an overall increase in the achievement scores of students with disabilities. The current study aimed to replicate and extend Hanushek et al.'s research using scores from the End-of-Grade Tests of Reading Comprehension and Mathematics for 396,828 students with and without disabilities who participated in special and/or general education programs over 5 academic years in the state of North Carolina. Results of multiple regression analyses revealed support for the effectiveness of special education programs. After controlling for school characteristics, cohort, regression to the mean, and accommodations use, entry into special education programs remained a significant predictor of students' gain scores on large-scale assessments of reading and mathematics. Students with disabilities made significantly larger gains when enrolled in special education programs than when enrolled in general education programs, and gains were found for most of the subgroups of special education students examined in the study. A discussion of the implications for practice, limitations of the study, and suggestions for future research are provided.

Estimating the Effectiveness of Special Education

Using Large-Scale Assessment Data

by
Katherine Anne Ewing

A dissertation submitted to the Graduate Faculty of
North Carolina State University
In partial fulfillment of the
Requirements for the degree of
Doctor of Philosophy

Psychology

Raleigh, North Carolina

2009

APPROVED BY:

| | |
|---|---|
| Dr. Ann Schulte<br>Committee Chair | Dr. Jeffery Braden |
| Dr. William Erchul | Dr. Susan Osborne |

# BIOGRAPHY

Katie was born in Indianapolis, Indiana on June 21, 1980 to Mark and Janet Raker. After graduation from Brebeuf Jesuit Preparatory School in 1998, she attended Miami University in Oxford, Ohio. In December 2001, she graduated Magna Cum Laude with a Bachelor of Arts and Departmental Honors in Psychology. Her Honors thesis, "Personality traits and social anxiety: Understanding comorbidity within a non-clinical, college population," was written under the direction of Dr. Mia Biran.

She began her graduate study in Psychology at Wake Forest University in Winston-Salem, North Carolina, from which she graduated Magna Cum Laude with a Masters of Arts in May 2004. She completed her Master's thesis, "Do parents matter? An examination of parental attitudes and involvement as predictors of children's social success within integrated and general education preschools," under the direction of Dr. Deborah Best. Katie then enrolled in the doctoral School Psychology program at North Carolina State University in Raleigh, North Carolina. She will complete her degree requirements in August 2009 and graduate Summa Cum Laude in December 2009.

Katie completed an internship with the NCSU Psychoeducational Clinic and the Lancaster County School District during the 2007-2008 school year. She currently lives in Charlotte, North Carolina and works as a school psychologist in Lancaster, South Carolina.

# ACKNOWLEDGMENTS

I want to thank my advisor, Dr. Ann Schulte, for the time and energy she has dedicated to my research project, as well as my professional development, over the past 5 years. Her guidance and support have helped me grow as a researcher, a writer, and a school psychologist. I would also like to thank Drs. Jeffery Braden, Atsushi Inoue, and Adam Meade for their input regarding my statistical procedures. Thank you to Drs. Jeffery Braden, William Erchul, and Susan Osborne for their feedback regarding my research and their participation in my preliminary examinations and my dissertation proposal and defense meetings.

I would like to thank Dr. Steven Rivkin for providing additional information about his 1998 and 2002 research with Drs. Hanushek and Kain, in order to help me replicate their procedures to the best of my abilities. I would like to thank the researchers at the North Carolina Education Research Data Center at the Center for Child and Family Policy at Duke University who made it possible for me to pursue this project by approving my research proposal and providing me with access to these data sets.

I would also like to thank my parents for their love and support, as well as their commitment to providing me with an excellent education. Finally, I would like to express my gratitude to my husband, James, whose love and friendship are more important to me than I can describe. I truly appreciate his unending patience and encouragement throughout my graduate career.

# TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

# CHAPTER ONE

## Introduction

Concern over the quality of education provided for children with disabilities has been present for centuries. However, a lack of sufficient evidence exists to determine whether current special education programs result in academic achievement gains for students with disabilities. The current paper will provide a review of previous research and legislation regarding the education of students with disabilities followed by the results of a study that utilized a new methodology for examining outcomes for such students.

Chapter Two provides a review of federal mandates and research concerning the education of students with disabilities. This review begins by examining how questions regarding the education of students with disabilities have shifted from ones of access to ones of effectiveness. In the early 1800s, children with disabilities were denied a formal education. However, over the next 150 years, the education of students with disabilities moved from occurring within separate schools to occurring within separate classrooms to occurring within the general education classroom whenever possible (Dunn, 1968; McDonnell, McLaughlin, & Morison, 1997). As their access to the general education curriculum increased over time, concerns began to grow regarding the quality of education provided to students with disabilities. This concern led to the passage of the original Education for All Handicapped Children Act in the 1970s. However, since its inception, the effectiveness of special education services provided to children through this legislative mandate has been questioned (e.g., Carlberg & Kavale, 1980).

Throughout the 1980s and 1990s, researchers have examined the outcomes for students with disabilities and tried to pinpoint what makes special education "special" (e.g., Fuchs & Fuchs, 1995). Some of this research examined the effectiveness of special education as a program, and other research specifically examined the effectiveness of different placement options (e.g., segregated classrooms, resource programs, full inclusion). Thus, the emphasis in special education research has changed from access to outcomes for students with disabilities.

This literature review then examines more recent changes in special education policy and research methodology during the standards-based reform movement. This movement began politically with the Goals 2000: Educate America Act, which posited that students with and without disabilities would demonstrate improved academic performance if they were expected to meet specific content and performance standards (McDonnell et al., 1997). Later, the 1997 reauthorization of the Individuals with Disabilities Act and the 2001 authorization of the No Child Left Behind Act more specifically required that (a) students with disabilities participate in large-scale assessments as a way to improve their academic achievement, and (b) states be held accountable for the achievement outcomes of all students, including those with disabilities. The use of large-scale assessments of academic achievement for both students with and without disabilities has provided researchers with a new outcome measure that can be used to examine the effectiveness of special education across schools, school districts, or states.

Next, an examination of the participation of students with and without disabilities in large-scale assessments is provided. Specifically, the consequences of large-scale assessments for students with and without disabilities, the inclusion of students with disabilities in accountability reports, the use of accommodations for students with disabilities, and the attitudes of teachers and parents regarding the inclusion of students with disabilities in large-scale assessments is summarized in this section. Additionally, the few research studies that have begun to use large-scale assessment data as outcome variables in special education research are briefly reviewed (e.g., Schulte & Villwock, 2004; Ysseldyke & Bielinski, 2002). Chapter Two concludes with a more detailed summary and critique of a study by Hanushek, Kain, and Rivkin (1998), which evaluated the effectiveness of special education programs in Texas by longitudinally analyzing students' gain scores on a large-scale assessment of student achievement.

Hanushek et al. (1998) conducted one of the first studies to use large-scale achievement test data to examine the effectiveness of special education placement for children with disabilities. They found that special education placement resulted in an overall increase in achievement scores for children served through its programs. This research was important because it used state-level data to examine the effectiveness of special education services and addressed long standing claims that national special education policy had resulted in services that were not only expensive, but ineffective (Fuchs & Fuchs, 1996). Additionally, it included a new methodology for examining the effectiveness of special education programs.

As is demonstrated in Chapters Three and Four, the current study aimed to replicate and extend Hanushek et al.'s (1998) research using large-scale assessment data for students with disabilities in North Carolina. Three years of End-of-Grade (EOG) Reading Comprehension and Mathematics Tests scores from 396,828 students from five cohorts were analyzed in order to estimate the effectiveness of special education programs in the state of North Carolina. The current study intended to increase the external validity of Hanushek and colleagues' results by replicating their findings within another state. It was also intended to extend their research by considering two additional factors that may have affected their results: (a) regression to the mean, and (b) the use of accommodations by students with disabilities during their participation in large-scale assessments.

The results of this study are presented in Chapter Five. Multiple regression analyses provided additional support for Hanushek and colleagues' (1998) findings that entering into special education programs improves the achievement scores of students with disabilities. Gains upon entrance into special education programs were found even after controlling for regression to the mean and the use of test accommodations. Chapter Six provides a discussion of these results within the context of previous research, explores the significance of the current study in the field of special education research, describes the limitations of the current study, and presents possible opportunities for future research endeavors.

# CHAPTER TWO

# A Review of the Literature

## *Access and Outcomes: An Historical Perspective on Special Education*

One of the most longstanding, controversial issues in education involves how to educate students with disabilities appropriately. Initially, these children were completely excluded from the formal education system. Over time, however, students with disabilities have moved from being educated within separate schools to being educated within separate classrooms to being educated within a general education classroom as much as possible. Throughout this progression, questions regarding the education of students with disabilities have shifted from ones of access to ones of effectiveness. The following section will (a) describe this progression for the education of students with disabilities, and (b) summarize classic research regarding the effectiveness of special education programs.

### *Moving From Exclusion From Schools to Inclusion Within the Classroom*

In the early 19th century, most children with severe mental and physical disabilities were removed from society and raised in asylums where they did not receive a formal education (McDonnell et al., 1997). During the mid-19th century, in response to this isolating, institutional approach, reformists such as Samuel Gridley Howe and Charles Loring Brace advocated for placing children with disabilities with families that could provide them with a more positive atmosphere in which to develop (Dorn, Fuchs, & Fuchs, 1996). Between 1852 and the end of World War I, most U.S. states enacted compulsory education laws for children; however, children with disabilities determined to be “uneducable” were still excluded from school (McDonnell et al.). Even when students with disabilities were granted access to a formal education, they continued to be segregated from students without disabilities and educated within self-contained, special schools. Students with disabilities have only routinely been provided with a formal education in separate, special classrooms since the mid-1900s (Dunn, 1968). Thus, from the early 1800s through the mid-1900s, the initial exclusion of students with disabilities from the formal education system was gradually replaced with access to education, first within segregated schools and later within segregated classrooms.

Beginning in the mid-1900s, many students with academic difficulties were placed in separate special education classes because it was believed the smaller class sizes, specially trained teachers, and specially designed curricula would improve their academic achievement (Kavale & Forness, 2000; Madden & Slavin, 1983). However, researchers quickly began to question the effectiveness of special education programs. Dunn (1968) was one of the first authors to address this issue when he argued that there was a lack of empirical evidence to support the placement of students with disabilities in segregated classrooms. Additionally, he cited research suggesting that segregating students with disabilities increases their feelings of inferiority and difficulties being accepted. Instead, he suggested that these students be educated in a more inclusive environment with their non-disabled peers. Similarly, Deno (1970) proposed a cascade system of special education services that was designed to allow educators to make individual decisions regarding the environment in which students with disabilities could be educated. This cascade system resembles the currently-used continuum of placement alternatives for students with disabilities that stemmed from legislative mandates in the 1970s.

Beginning in the early 1970s, a variety of legislative mandates changed the way in which students with disabilities were educated. First, two important federal court cases ruled against school districts for continuing to exclude students with disabilities. In 1971, in *Pennsylvania Association of Retarded Children (PARC) v. Commonwealth*, the court ruled that schools had the responsibility to enroll children with mental retardation. In 1972, in *Mills v. Board of Education*, the court ruled that a school could not exclude students with disabilities due to a lack of funding (Gartner & Lipsky, 1987; Huefner, 2000). Then, the Rehabilitation Act of 1973 prohibited public and private organizations receiving federal funding from discriminating against individuals with disabilities. This act stemmed from the Civil Rights movement and had implications for all persons with disabilities; however, it specifically prohibited schools from discriminating against students with disabilities (Schulte, Osborne, & Erchul, 1998). These initial mandates aimed to include higher numbers of students with disabilities in the formal education system; however, in 1972, only 40% of students with disabilities were being provided educational services (McDonnell et al., 1997).

In response to this continued exclusion, in 1975, Congress passed the Education for All Handicapped Children Act, PL 94-142 [later renamed the Individuals with Disabilities Education Act (IDEA) and now known as the Individuals with Disabilities Education Improvement Act (IDEIA)]. PL 94-142 mandated that students with disabilities be given a free appropriate public education (FAPE) based on an Individualized Education Plan (IEP). It also supported the earlier notion proposed by Dunn (1968) and Deno (1970) that students with and without disabilities should be educated together. Specifically, students with disabilities were required to be educated within the least restrictive environment (LRE) possible among a continuum of placement alternatives (Huefner, 2000; Kavale & Forness, 2000; Schulte et al., 1998). The LRE mandate led to the mainstreaming of many students with disabilities in general education classrooms with their non-disabled peers. According to Osborne and Dimattia (1994), relevant court decisions following the original IDEA mandate “clearly indicate that an inclusionary placement must be the placement of choice and that a student with disabilities may be excluded from a general education setting only in the face of strong evidence that the student cannot be satisfactorily educated in that setting” (p. 12).

The aim of the original IDEA legislation was to improve educational outcomes for students with more severe disabilities. However, according to Kavale and Forness (1999, 2000), as the law was implemented, it gradually became clear that the majority of students served in special education programs were those with mild disabilities (e.g., learning disabilities). These students tended to receive special education services through a resource model. This model involves educating many students with mild disabilities primarily within the general education classroom, although special education teachers may remove them from this classroom for specific periods to provide supportive, individualized academic instruction. As the numbers of students served began to rise, research regarding the effectiveness of special education services in a range of settings (e.g., self-contained classrooms, resource rooms) became increasingly important. Indeed, much of the focus in special education policy discussions shifted from ensuring that students with severe disabilities had access to a formal education to examining the outcomes for students with a wide range of disabilities within special education programs.

*Classic Special Education Outcomes Research*

The original passage of the IDEA changed the way in which students with disabilities were educated and increased the desire for researchers to examine the effectiveness of special education programs. Many of these studies involved relatively small sample sizes and examined outcomes for students in general education classrooms versus a variety of special education settings. Additionally, much of this research was methodologically flawed. Three of the most common flaws included (a) comparing matched groups of students in general and special education classrooms without using random selection, (b) inadequately describing the instruction provided to students in the control and treatment groups, and (c) using grade-equivalent scores from standardized tests to determine program efficacy (Kavale & Forness, 1999; Madden & Slavin, 1983; Zigmond, 2003). Among the methodologically sound studies, results tended to suggest that students with disabilities who participated in mainstream programs had higher academic and social outcomes than did students with disabilities who participated in various forms of special education; however, these results were not consistent (Madden & Slavin, 1983; Zigmond, 2003). Additionally, some research demonstrated that although special education programs could improve the achievement of students with disabilities in an absolute sense, these programs could not sufficiently close the achievement gap between students with and without disabilities (Shinn, 1986).

Carlberg and Kavale’s (1980) meta-analysis was one of the first studies to provide a comprehensive examination of the outcomes for students with disabilities in a variety of education settings. Their meta-analysis revealed that placement in a special class had an overall effect size of -0.12. This finding resulted in special class placement being listed as one of five “interventions that don’t work,” (along with social skills training, modality instruction, the Feingold Diet, and perceptual training) in a later publication (Forness, Kavale, Blum, & Lloyd, 1997). In more specific analyses regarding the placement of students with different types of disabilities, Carlberg and Kavale found that the effect size of special class placement for students with more severe disabilities was larger than the effect size of special class placement for all students with disabilities combined. For example, the effect size for special class placement for students with mild mental retardation (i.e., they had a standardized IQ score between 50 and 75) was -0.14 and for students considered to be slow learners (i.e., they had a standardized IQ score between 75 and 90), -0.34. However, for students with less severe cognitive disabilities (i.e., learning disabilities or behavioral/emotional disorders), special education placement was more beneficial (effect size = 0.29) than placement within the general education classroom. Overall, this complex meta-analysis demonstrated that, on average, groups of students with a variety of disabilities do not benefit from special education placement. However, students with mild disabilities were more likely to benefit from special education programs than were students with more severe cognitive difficulties. Recently, McLeskey (2004) identified the 50 most frequently cited articles in special education between 1960 and 1996 and labeled them “classic” articles that have shaped the field. At the time of McLeskey’s publication, Carlberg and Kavale’s meta-analysis had already been cited over 100 times, making it one of the most influential, classic studies examining outcomes for students with disabilities.

Carlberg and Kavale (1980) statistically analyzed the effectiveness of special education through a meta-analysis. In contrast, Sindelar and Deno (1978) conducted a literature review to examine the efficacy of resource placement for students with disabilities in terms of both academic and social outcomes. Their review of 11 studies that used comparison groups to assess academic outcomes revealed the following tentative conclusions: (a) for students with mental disabilities, the efficacy of resource placement is unclear; (b) for students with less severe disabilities, resource placement is generally more favorable than general education placement without resource support; and (c) additional research is needed to assess the efficacy of resource placement. A few years later, Leinhardt and Pallay’s (1982) literature review examined the effectiveness of many different special education settings (e.g., self-contained classrooms, resource rooms, general education classrooms, indirect services). They concluded that positive academic outcomes for students with disabilities can occur regardless of their educational placement. Thus, children with disabilities should be educated within the LRE and educators ought to focus on understanding the *variables* related to success for students with disabilities, rather than on the *place* in which education occurs.

Similarly, Maddin and Slavin (1983) reviewed previous studies examining both academic and social outcomes for students with mild handicaps. They concluded that much of the early research conducted during the 1950s and 1960s regarding the effectiveness of special education in terms of academic outcomes (which demonstrated no difference between placements) was not methodologically valid. However, even the later, methodologically sound studies generated mixed results regarding academic outcomes. Some revealed that general education was more beneficial than special education; others showed no differences between programs. These authors drew the following tentative conclusions regarding academic achievement outcomes for students with mild disabilities: (a) when individualized instruction is utilized within the general education classroom, the achievement of students with disabilities is higher than when they are educated within special education classes; (b) the effectiveness of general education class placement depends on the alterations to the program in that classroom; and (c) the evidence regarding the effectiveness of part-time resource room placement is unclear.

In addition to concerns regarding the ability of special education programs to improve the academic achievement of students with disabilities, concern also existed regarding the ability of such programs to decrease the achievement gap between students with and without disabilities. Shinn (1986) found that students with mild handicaps in grades one through six demonstrated absolute growth on curriculum-based assessments of reading skills (i.e., the number of words read correctly) between fall, winter, and spring assessments. However, the achievement gap between students in special and general education was not reduced over this same time period. In fact, for students in many of the upper grades, the discrepancy grew larger. Thus, special education classes may improve the achievement of students with disabilities, but not at a rate that allows them to close the gap between themselves and their non-disabled peers. When considered together, the classic special education studies reveal mixed results regarding the effectiveness of special education programs. Shinn considered these findings to indicate a need for policy change: “Without intervention effectiveness data or with data that are indicative of marginally effective or ineffective programs, concern about the entire classification process is justified” (p. 50).

Because of the lack of conclusive empirical support for the use of segregated special education settings, some educators pushed for a more inclusive placement for students with disabilities in the general education classroom. This movement occurred during the mid- to late-1980s and was called the Regular Education Initiative (REI). The major goal of the REI involved merging general and special education into one system with the hope that a significantly increased number of students with mild disabilities would be moved into the general education classroom (Fuchs & Fuchs, 1994; Will, 1986). This unified system would allow for the individual differences of all students to be addressed in the general education classroom without the need for the classification and labeling of students with disabilities (Reynolds, Wang, & Walberg, 1987; Stainback & Stainback, 1984). According to Gartner and Lipsky (1987), the REI received support because a review of the literature “found full- or part-time regular class placements more beneficial for students’ achievements, self-esteem, behavior, and emotional adjustment”

(p. 375). Additionally, full inclusion was believed to be advantageous because: (a) the use of classwide accommodations would improve the achievement of students with and without disabilities, (b) a decrease in the stigma associated with disabilities would occur, and (c) students with disabilities would not miss out on general education instruction or social interactions that occur when they are pulled out for resource room instruction (Zigmond & Baker, 1996). Despite these potential advantages, the REI had limited empirical support (Kavale & Forness, 2000). Zigmond and Baker also reported that the main disadvantage to full inclusion is that students with disabilities do not receive direct and intense intervention when they are solely educated within the general education classroom.

Hocutt's (1996) more recent review of the literature led her to conclude her article with a statement that echoed the earlier findings of Maddin and Slavin (1983): "Research supports the continuation of efforts to improve academic and social outcomes for students with disabilities in both special and general education settings and indicates that instruction, not setting, is the key to achievement of success as measured by student outcomes" (p. 98). Additionally, until schools have adequate resources to assist all students with disabilities in the general education classroom, there is a need for special education programs to continue to exist. Finally, Hocutt stated that utilizing effective instructional methods such as direct instruction, cooperative learning, peer tutoring, cognitive strategy instruction, and transenvironmental programming in both special and general education programs may be the most effective way to improve all students' academic achievement. In sum, the classic research regarding the effectiveness of special education reveals mixed results and questions regarding outcomes for students with disabilities remain.

*The "Special" in Special Education*

As an alternative way of assessing the effectiveness of special and general education programs for students with disabilities, some researchers have attempted to identify the variables that make special education "special." Fuchs and Fuchs (1995) argued that special education programs have unique resources that general education programs cannot provide. For example, special education involves the use of IEPs (which allow teachers to create and attain specific short- and long-term goals for each student), smaller class sizes (which allow for more individualized instruction), and more highly trained teachers (who tend to have more advanced degrees than general education teachers). These authors highlighted the previous research findings that special education is more effective than general education for students with certain types of disabilities (i.e., those with milder disabilities).

In the same year, Scruggs and Mastropieri (1995) evaluated the effectiveness of full inclusion programs using a set of variables they asserted are essential for successful special education. The four variables they identified as essential for successful special education can be summarized with the acronym PASS. The PASS variables include: (a) *p*rioritizing objectives within the curriculum and allocating resources accordingly; (b) *a*dapting materials, instructional strategies, and instructional environments to meet students' needs; (c) utilizing important teacher presentation variables with the *S*CREAM method (structure, clarity, redundancy, enthusiasm, appropriate pace, and maximized engagement); and (d) systematically monitoring students' progress. They claimed that the five full inclusion case studies described by Zigmond and Baker (1995) in the same journal issue did not fulfill the PASS criteria. Thus, although Scruggs and Mastropieri stated that they support the education of students with disabilities in the least restrictive environment, they did not believe that these particular full inclusion programs were sufficiently "special" to meet the needs of students with learning disabilities.

*Preparing for Change*

As the previous summary demonstrated, over time, questions regarding the education of students with disabilities have shifted from ones of access to ones of outcomes. Students who were once ignored and excluded have gradually become the focus of a great deal of research. Although studies examining the effectiveness of special education have been conducted for over half a century, the earliest studies were often methodologically flawed. With the passage of the original IDEA, researchers became increasingly interested in determining whether special education programs were producing improvements in academic achievement for students with disabilities. The individual studies, literature reviews, and meta-analyses during the 1980s and early 1990s produced mixed results regarding the outcomes for students with disabilities educated in special education settings. However, these studies demonstrated that students with disabilities consistently perform at a level that is below that of their peers, regardless of where they are placed. For example, Schulte (1996) reminded readers that, "The longitudinal outcomes for students with learning disabilities, whether in inclusive or restrictive settings, are not positive. At present we do not have programs for students with learning disabilities that have demonstrated track records of producing high levels of literacy, numeracy, and content knowledge” (p. 204). Despite their varied methodologies and discouraging results, classic special education studies did provide a starting point for examining the effectiveness of special education programs, and their results underscore the need for more research about the impact of special education programming and the factors consistent with effective programs.

As the interest in outcomes for students with disabilities has increased, so has the interest in examining the cost of special education programs. According to a retrospective study by Chaikind, Danielson, and Brauen (1993), the cost of educating a student in a special education program is approximately 2.3 times as much as the cost of educating a student in a general education program. This ratio has remained relatively stable over time: from the 1970s through the 1990s, the total expenditure for students receiving special education services has remained approximately twice the cost for students in the general education program. A more recent examination of special education expenditure by Chambers, Parrish, and Harr (2002) revealed that the total amount spent providing both special and general education services to students with disabilities was $77.3 billion in the United States during the 1999-2000 academic year. This figure represents an average of $12,474 spent to educate each student with a disability. Meanwhile, an average of $6,556 was spent to educate each student without a disability in the general education program, indicating that the additional expense of educating a student with a disability in this country was $5,918 that year. Expensive special education programs have come under increased scrutiny as the number of students receiving such services continues to grow.

According to the National Center for Education Statistics (2006), the number of students receiving special education services in the United States has increased from almost 3.7 million in the 1976-1977 school year to over 6.6 million students in the 2003-2004 school year (see Figure 1). These numbers reflect an increase from 8.3% to 13.7% of the total student population (see Figure 2). Additionally, the composition of students with disabilities has changed over the course of the past 30 years. Although the number and proportion of students with more severe disabilities [e.g., mental retardation (MR)] has decreased, the number and proportion of students with less severe disabilities [e.g., specific learning disabilities (SLD), behavioral-emotional disabilities (BED)] has increased (see Figure 3). Because of the added expense associated with providing special education services, as well as the growing number of students receiving these services, it is important to assess the outcomes for students with disabilities. Additionally, as increasing numbers of students with mild disabilities are identified, a larger proportion of the population of students with disabilities is being educated primarily within the general education classroom. Thus, there is a need to examine the effectiveness of special education programs using a new set of outcomes.

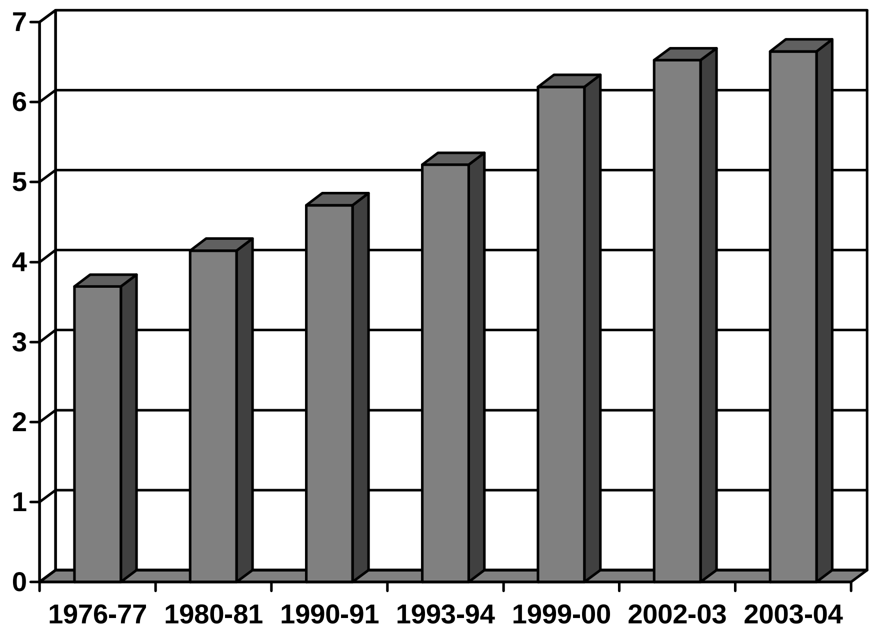


Figure 1. The number (in millions) of students receiving special education services in the U.S. (National Center for Educational Statistics, 2006).

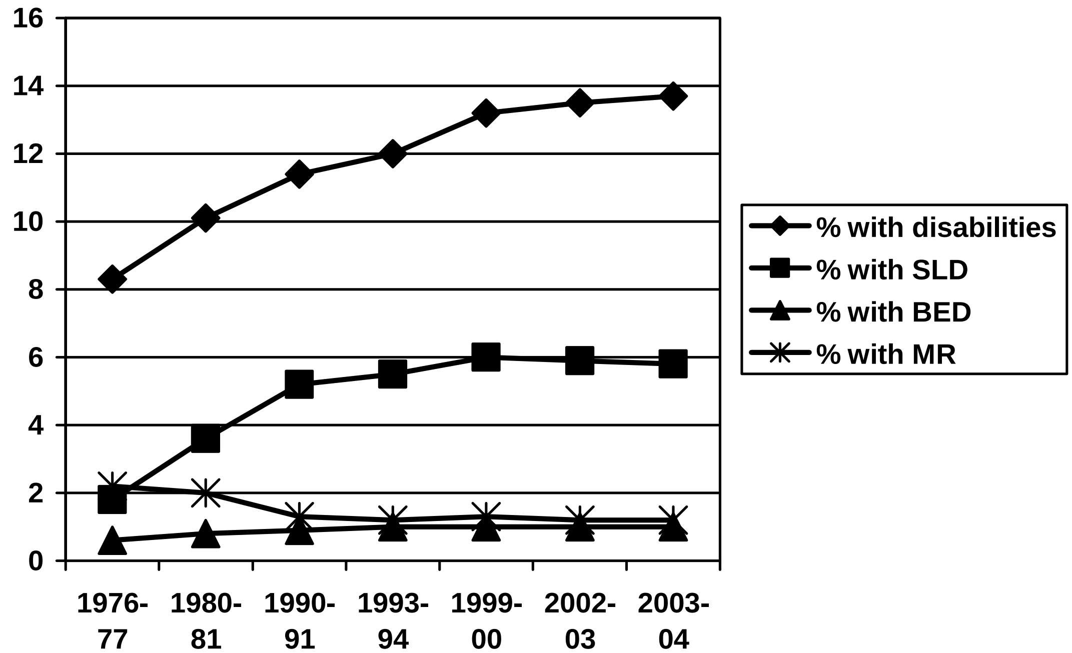


Figure 2. The proportion of students receiving special education services in the U.S. (National Center for Educational Statistics, 2006).

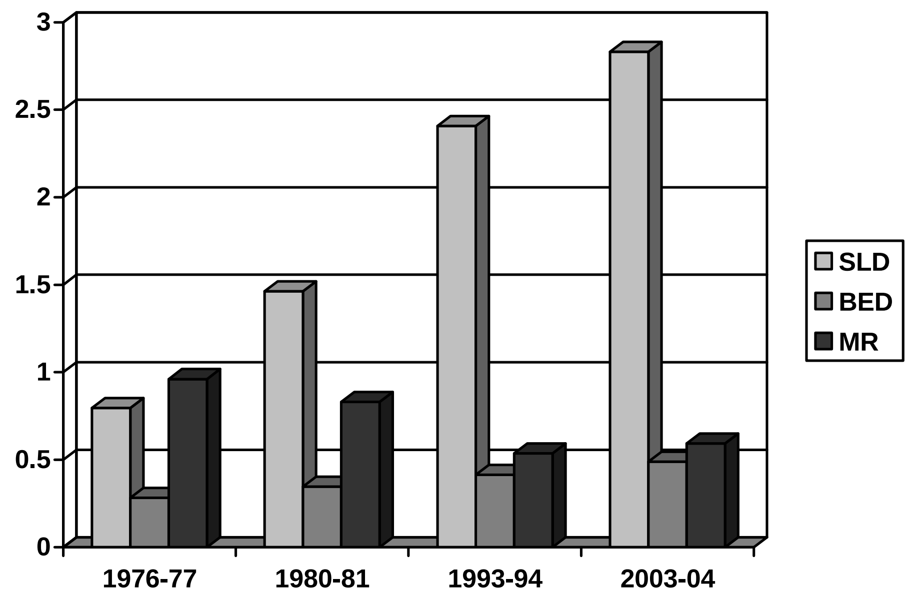


Figure 3. The number (in millions) of students receiving special education services in the U.S., by disability type (National Center for Educational Statistics, 2006).

---

Kavale and Forness (1999) summarized the history of special education research with the following statement: “The emphasis brought about by the cascade model has made it the focus of attention; yet, there was very little advantage to any placement, as shown by the rather small ES associated with evaluations of different settings. Two explanations are possible: special education may indeed be ineffective, and the methodology for measuring outcomes may not be appropriate” (p. 1007). Indeed, the effectiveness of special education must now be estimated by measuring a different set of outcomes. With the passage of new legislation in the late 1990s and early 21st century, students with disabilities have been given access to standards in the general education curriculum. Now, educators have the opportunity to change the way special education effectiveness research is conducted. Specifically, the standards-based reform movement allows for the examination of outcomes for students with disabilities through their scores on large-scale assessments. Because scores on large-scale assessments are examined in the current study, the next section of this paper will describe how the standards-based reform movement led to the wide-spread use of large-scale assessments for both students with and without disabilities.

### *Standards-Based Reform and its Implications for Special Education*

The standards-based reform movement began in the mid-1990s as a way to increase students' academic achievement by providing them with well-defined, high standards for academic success. Once these standards were in place, students, schools, and even states became accountable for ensuring that students were making yearly academic progress in meeting these standards. Today, large-scale assessments are given to students with and without disabilities on an annual basis to examine their achievement in the areas of reading and mathematics. Relatively little research has examined the use of large-scale assessments for students with and without disabilities. However, research regarding the consequences of these assessments, the inclusion policies for students with disabilities, the use of accommodations by students with disabilities, and the attitudes of parents and teachers toward the inclusion of students with disabilities in large-scale assessments has been completed. Additionally, the use of large-scale assessments of academic achievement for both students with and without disabilities has provided researchers with a new outcome measure that can be used to examine the effectiveness of special education programs.

*The Standards-Based Reform Movement*

In 1994, the Goals 2000: Educate America Act was passed to encourage individual states to adopt content and performance standards for all students. Content standards were defined as broad descriptions of the knowledge and skills students ought to acquire in each content area (e.g., reading, mathematics) and performance standards were defined as specific examples of what students must be able to do as a demonstration of proficiency in these areas. This federal legislation provided momentum for the standards-based reform movement, which posited that all students can meet high academic standards when those standards are clearly defined and taught by teachers. In other words, having common standards serves as a means for improving student outcomes. Although there is a lack of research to demonstrate that focusing on standards will directly improve student outcomes, research has revealed a relationship between the type and difficulty of courses taken and achievement outcomes for students. Thus, standards-based reform policies focused on increasing the content and performance standards students must attain at each grade level in order to improve their achievement (McDonnell et al., 1997).

The standards-based reform movement had five underlying assumptions: (a) standards are applicable to all students (including students with disabilities), (b) content and performance standards can be transformed adequately into curriculum objectives, (c) students’ performance can be measured reliably and validly (primarily by assuring the content validity of outcome measures), (d) individual schools and classrooms can include instruction that is consistent with the standards, and (e) standards-based reform is only one component of a strategy for improving schools (McDonnell et al., 1997). McLaughlin and Warren (1992) argued that students' outcomes ultimately ought to be used to improve instruction and, thus, improve the educational process. For this reason, outcomes for all students (including those with disabilities) must be measurable in such a way that allows for comparisons to be made among schools, school districts, and states. Recently, student outcomes have begun to be measured through scores on large-scale assessments, which are designed to measure students' performance as the extent to which they have mastered content standards at a particular grade level (McDonnell et al.).

*Moving From Inclusion in the Classroom to Inclusion in Large-Scale Assessments*

When the original IDEA was passed in 1975, it ensured that students with disabilities were educated with their non-disabled peers to the greatest extent possible. Despite their increased inclusion in the general education classroom, students with disabilities did not traditionally participate in large-scale assessments of academic achievement due to concerns regarding the consequences of making modifications to the tests. In fact, it has been suggested that low-achieving children were either retained in their grade or referred to special education programs as a way to ensure their exclusion from large-scale accountability assessments. This phenomenon, sometimes termed "academic red-shirting," resulted from incentives being given to schools placing more low-achieving students in special education programs, thereby improving the achievement scores of the students in general education classrooms (Elliott, Erickson, Thurlow, & Shriner, 2000; Zlatos, 1994).

As the standards-based reform movement gained speed, it became clear that students with disabilities ought to be included in accountability practices. Thus, academic red-shirting practices diminished with the reauthorization of the IDEA in 1997 (as PL 105-17), which stated that students in special education programs must be allowed to participate in large-scale assessments (or in alternative assessments) with appropriate accommodations (Huefner, 2000; Schulte et al., 1998). However, the decision to include an individual student with disabilities in such assessments is made at the school level, usually by the professionals involved in the child's IEP team (Koretz & Baron, 2003). A study by the National Center on Educational Outcomes found that across different states, factors such as the course content, the input of the parent/guardian, and the percent of time a student spends receiving special education services are also considered when making the decision of whether a student with disabilities participates in large-scale assessments (Thurlow, House, Scott, & Ysseldyke, 2000). Although federal legislation does not specify what proportion of students with or without disabilities must participate in large-scale assessments, most states assume that 98% of all students should participate. According to Braden (2002), "given that most states classify 12% of the students in a given grade as disabled, it means that approximately 85% of students with disabilities must be included in state systems to meet the overall 98% participation goal" (p. 310). Thus, the 1997 reauthorization of IDEA resulted in significantly increased numbers of students with disabilities participating in large-scale achievement tests.

More recently, the No Child Left Behind Act (NCLB; 2001) increased accountability for students in federally funded elementary and secondary schools by requiring all children in the third through eighth grades to be tested annually in both reading and math. This legislation mandated that each state implement an accountability system to ensure schools make adequate yearly progress (AYP) in improving the achievement of all students, with the goal that all groups of students reach proficiency by the year 2014. States must not only demonstrate improvement in overall achievement, but must also exhibit improved achievement for students in various disaggregated groups (including students with disabilities) to ensure that no group or individual child is left behind.

These large-scale assessments are often referred to as “high-stakes tests” because of the consequences states associate with students’ scores at both the school- and student-level. School-level rewards can include additional funding or recognition; school-level sanctions can include decreased autonomy, probation, loss of accreditation, and funding loss. At the student-level, the results of large-scale assessments may be used to make grade promotion or graduation decisions (Jones, 1997; McDonnell et al., 1997). For example, within the state of North Carolina, teachers can receive bonuses if students’ average growth meets or surpasses the standards set for their grade level. Likewise, schools that meet or exceed AYP goals may be recipients of State Academic Achievement Awards. On the other hand, a school may be taken over by a state assistance team if fewer than 50% of their students cannot meet these growth standards.

Additionally, test scores may be reported in the local newspaper by school, grade level, or classroom (Mehrens, 1988; Schulte, Villwock, Whichard, & Stallings, 2001).

*The Use of Large-Scale Assessments for Students With and Without Disabilities*

To date, relatively little research has examined the use of large-scale assessments for students without disabilities; even less research has examined the use of large-scale assessments for students with disabilities. However, the research that has been conducted has covered a wide variety of topics. This section will summarize previous research regarding the intended and unintended outcomes of high-stakes assessments for students with and without disabilities, state-level accountability for students with disabilities, the use of accommodations on large-scale assessments for students with disabilities, and teacher and parent attitudes regarding testing and accommodating students with disabilities.

*Intended and unintended outcomes of high-stakes assessments.* The NCLB Act called for the use of these high-stakes tests with the expectation that increasing knowledge about students' achievement, rewarding those who make achievement gains, and punishing those who do not make adequate gains would motivate both teachers and students and result in improved student outcomes (Roach & Frank, 2007). However, more research is necessary to determine the effects of high-stakes assessments on students' achievement, as "the findings from the most rigorous studies on high-stakes testing do not provide convincing evidence that high-stakes testing has the intended effect of increasing student learning" (Nichols, 2007, p. 48).

Many positive outcomes of large-scale assessments have been hypothesized. The positive, intended outcomes for students, teachers, and schools include the following: increased passing rates, increased number of students receiving a diploma, increased academic expectations, increased focus on achievement in IEPs, improved instruction, increased use of accommodations for students with disabilities, and improved ability of parents to communicate with educators regarding achievement (Ysseldyke et al., 2004). Braden (2002) noted that other intended outcomes include improved curricular alignment and increased motivation in both teachers and students. Research by Christenson, Decker, Triezenberg, Ysseldyke, and Reschly (2007) reveals that special education teachers, general education teachers, and school psychologists believe that high-stakes assessments have led to increased monitoring of student progress and increased efforts to improve the performance of struggling students in the classroom.

Despite the positive, intended outcomes of high-stakes assessments, negative outcomes may exist as well. Braden (2002) listed five potential unintended consequences of these assessments: curricular narrowing (i.e., educators “teach to the test”), scholastic demoralization (e.g., students who experience repeated poor performance experience frustration, which they then generalize to other academic tasks), corruption (e.g., cheating), inequitable educational opportunity (e.g., tracking of low scoring students), and panic (i.e., high levels of anxiety among students, teachers, and administrators). Research by Ysseldyke and colleagues (2004) revealed that special education teachers, administrators, and parents of students with disabilities have reported concern regarding anxiety as a possible unintended outcome for students with disabilities, in particular.

Although studies have not specifically assessed anxiety as an outcome for students with and without disabilities, concern exists that such anxiety might lead to other negative consequences (e.g., increased cheating).

A review of the research regarding the impact of high-stakes testing on students, parents, teachers, and schools is beyond the scope of this literature review. However, Mehrens' (1988) comprehensive review revealed that high-stakes assessments can have negative consequences for teachers (e.g., increased stress and burnout, decreased morale, increased probability of unethical behavior) and students (e.g., increased feelings of inadequacy, decreased motivation, decreased morale). A more recent review by Jones (2007) found that the consequences of high-stakes assessments affect teachers' and students' motivation, as well as teachers' instructional methods, in negative ways. Thus, the use of large-scale assessments can potentially result in both positive and negative outcomes for students, teachers, and schools.

Together, the reauthorization of the IDEA in 1997 and the passage of the NCLB Act in 2001 require that (a) students with disabilities be included in large-scale academic assessments and (b) schools be responsible for demonstrating that students in both general and special education programs are making annual progress on these high-stakes, large-scale assessments. Although more current research is needed to identify the positive and negative consequences of high-stakes tests (Ysseldyke et al., 2004), these statewide tests are one of the best measures of academic achievement currently available for individuals with and without disabilities across schools.

*State-level accountability for students with disabilities.* Large-scale assessments of academic achievement are not only used to assess the performance of individual students, but also to assess the overall performance across schools, school districts, and states. As mandates from the 1997 reauthorization of the IDEA and the NCLB Act have been put into practice, changes in state-level accountability practices have occurred. Elliott et al. (2000) examined these changes over a 5-year period. Their findings revealed that during the mid-1990s, states demonstrated that they were increasingly including students with disabilities in their accountability practices. For example, the number of states that included school participation (e.g., attendance, placement, suspension, or expulsion information), exit information (e.g., dropout/graduation rate, dropout reason, diploma type), and achievement data on students with disabilities in state-level accountability systems was much greater in 1995 than in 1991. Specifically, only 4 states included information about the school participation of students with disabilities in their accountability systems in 1991, whereas 26 states included this information in 1995. Similarly, the number of states providing exit information increased from 2 states in 1991 to 20 states in 1995; the number of states providing achievement data increased from 7 states in 1991 to 36 states in 1995. Unfortunately, during this same time period, there was no difference in the number of states that could provide a participation rate for students with disabilities on large-scale assessments. At both points in time, only 19 states reported that just less than half of the students with disabilities participated in large-scale assessments; other states were unable to provide an estimate of their participation.

*The use of accommodations on large-scale assessments.* When students with disabilities participate in large-scale assessments, they are permitted to use accommodations, as is deemed appropriate by an IEP team. Accommodations involve changes to the way a test is administered or to the way a student responds to a test; they should not interfere with the assessment of target skills (the skills the test is purporting to measure). Rather, accommodations are implemented to improve students' access skills, allowing them to appropriately demonstrate their target skills (e.g., being able to read the test, being able to complete the test within the allotted time frame). Accommodations differ from modifications, which alter target skills and, thus, change the content of the test (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Niebling & Elliott, 2005). Providing test accommodations to students with disabilities is thought to improve the validity of their scores by removing the irrelevant interference of their disabilities and allowing them to display their knowledge more accurately (Koretz & Barton, 2003; Niebling & Elliott, 2005).

Test accommodations can be classified into four types. The first type of accommodation involves changes to the *administration or presentation format* and includes having the test read aloud, using a Braille version of the test, or using a test with large print. The second type of accommodation involves changes to the *test's response format* and includes allowing students to provide oral, signed, typed, or non-verbally indicated responses. The third type of accommodation involves changes to the *setting of the test* (e.g., isolated setting, small group setting). Finally, accommodations can also be made to the *timing of the test*, such as providing extended time or additional breaks during the testing session (Braden, 2002; Thurlow et al., 2000).

The National Center on Educational Outcomes studied the use of accommodations across the United States in 1993, 1995, and 2000. In the most recent study, Thurlow et al. (2000) examined the written policies regarding the use of accommodations from all 40 of the 50 states with active policies. Eighty percent of these active policies allowed some accommodations of nearly every type. The eight most commonly provided accommodations across the country included the use of large print, Braille, a proctor/scribe, sign language, small group administration, extended time, individual administration, and having the test read aloud. Thus, state policies most frequently allowed for accommodations to the presentation format of a test. However, acceptable accommodations varied by state, making comparisons across states more difficult.

According to a literature review by Niebling and Elliott (2005), research regarding the use of testing accommodations has supported the following findings: (a) testing accommodations are frequently used together, rather than individually, (b) single-case designs are useful in evaluating the effects of accommodations, (c) students with disabilities can serve as their own comparison group when evaluating the effects of accommodations, (d) accommodations tend to increase the performance of students with disabilities by one-fourth to one-half of a standard deviation, and (e) accommodations also tend to increase the performance of students without disabilities, though to a lesser degree.

One of the most controversial accommodations is reading a math test to students with reading disabilities. Johnson (2000) provided this accommodation to students with and without disabilities on a large-scale mathematics achievement test to assess its validity. Results suggested that the accommodation of reading math items to students may have a differential effect for students with and without disabilities. For students with reading disabilities, the mean score on the Washington Assessment of Student Learning (WASL) increased by nearly 20 points when the test was read to them; for students without disabilities, the mean score on the WASL decreased by less than 2 points when the test was read to them. However, because this study was conducted with a small sample size (38 students per group), this effect approached, but did not quite reach significance ($p < .076$). Despite the limitations of this study (e.g., the small sample size, the removal of the "high stakes" associated with the assessment by telling students the examination was "didn't count" or was "for research purposes"), it provides some initial support for the validity of a specific accommodation for students with reading disabilities.

A study by Elliott, Kratochwill, and McKevitt (2001) examined the effects of accommodations on achievement test scores for students with and without disabilities. Math and science achievement tests were taken by both groups of students with and without the use of accommodations as a way to simulate large-scale assessments. They found that students with disabilities performed almost one standard deviation lower when they took the test without accommodations than when they took the test with accommodations. Also, the performance of students with and without disabilities was very similar when students with disabilities were given testing accommodations and students without disabilities were not given such accommodations. Finally, students without disabilities benefited somewhat from the use of accommodations. Although over 75% of students with disabilities demonstrated a medium to large effect size when using accommodations, the same was true for approximately half of the students without disabilities. Additionally, the use of accommodations led to negative effects for 17% of the students with disabilities and 7% of the students without disabilities. Thus, the authors concluded that accommodations tend to have a positive impact on the test scores of students with disabilities; however, accommodations affect the achievement scores of individual students differently.

More recently, Fletcher, Francis, Boudousquie, Copeland, Young, Kalinowski et al. (2006), examined the effects of accommodations on a high-stakes reading assessment, the Texas Assessment of Knowledge and Skills (TAKS), for students with and without disabilities. Students identified as having a reading disability (and below average decoding skills) and students with average decoding skills were either given no accommodations or a package of three accommodations specifically designed to reduce the impact of students' decoding difficulties on their TAKS scores. The accommodations included: (a) having proper nouns read aloud to them, (b) having the stems and possible responses of test items read aloud to them (after the students independently read the reading passage), and (c) taking the test in two sessions instead of one. Students with a reading disability performed better on the TAKS when they received these accommodations than when they did not. In fact, students with disabilities who received the accommodations were approximately seven times more likely to pass the TAKS than were students with disabilities who did not receive accommodations. In addition, the scores of students without reading disabilities were not significantly different for the students who were given accommodations and the students who were not. Again, accommodations were shown to improve the scores of students with disabilities, but to have no significant change on the scores of students without disabilities.

Lang et al. (2005) examined the reactions of students, teachers, and parents to the use of testing accommodations. The majority of students with and without disabilities reported that they felt more comfortable when the teacher provided testing accommodations (46% and 43.4%, respectively) than when the teacher did not (9.5% and 17.5%, respectively). However, a large proportion of students with (44.4%) and without (39.2%) disabilities reported no preference between conditions in terms of their comfort level. Additionally, the majority of students with (62.2%) and without (50%) disabilities reported that the test seemed easier when the teacher provided them with accommodations; a much smaller proportion reported that the test actually seemed easier without accommodations (9.4% of students with disabilities and 16.9% of students without disabilities). Finally, students with disabilities were more likely than their non-disabled peers to report that when using accommodations, they were better able to show their knowledge in both math and reading. In open-ended questions, students stated that they appreciated the use of accommodations, but did not always find them to be necessary. Parents and teachers tended to agree that the use of accommodations was a fair way to assess students with disabilities, as it makes their scores more comparable to the scores of non-disabled peers who do not use accommodations.

Although students, parents, and teachers recognize that accommodations can improve the test scores for students with disabilities, federal guidelines do not yet specify which accommodations are acceptable for use and which accommodations provide an unfair advantage for students with disabilities (Koretz & Barton, 2003). Until more specific direction is provided, IEP teams can use tools such as the *Assessment Accommodations Checklist* or *The Dynamic Assessment of Test Accommodations* to help determine if a test accommodation is appropriate (Niebling & Elliott, 2005). Additionally, Braden (2002) provided the following “best practice” guidelines regarding the inclusion of students with disabilities in large-scale assessments and the use of testing accommodations during these assessments: (a) students who do not participate in the general education curriculum probably should not participate in large-scale assessments; (b) when in doubt, allow students to participate in large-scale assessments because there is no evidence yet that they are harmed by their inclusion; (c) accommodations should already be provided in the student’s classroom for at least 4 to 6 weeks before they are used during a large-scale assessment; (d) accommodations should only address students’ access skills and should not alter target skills; (e) prior testing should be used as a guide when considering the use of accommodations; and (f) accommodations used and students’ responses to them should be documented for future reference.

*Attitudes regarding the use of large-scale assessments for students with disabilities.* Teachers and district-level personnel have been found to be “cautiously pessimistic" about the inclusion of students with disabilities in large-scale assessments for accountability purposes, as well as their use of accommodations on such assessments. Oregon teachers participating in focus groups stated concerns regarding the use of large-scale assessment data for accountability purposes of any kind. However, they were particularly concerned about the inclusion of students with disabilities in such assessments because of (a) the amount of testing students with disabilities must already endure, (b) the distress students with disabilities may feel in response to these tests, and (c) the pressure they, as teachers, feel because of the high stakes associated with the use of the results in making accountability decisions (Crawford, Almond, Tindal, & Hollenbeck, 2002). Through written responses and interviews, district-level personnel (e.g., teachers, administrators, school psychologists, and assessment coordinators) in Illinois also revealed concerns regarding partial participation by students with disabilities and the use of accommodations by students with disabilities (Shriner & DeStefano, 2001). Despite this apprehension by teachers and administrators, equity advocates argue that when students with disabilities are excluded from accountability programs, school personnel are more likely to ignore their needs, place them in low quality environments, deny them access to the general curriculum, and continue to have low expectations for their academic success (Braden, 2002; Zlatos, 1994). Although the inclusion of students with disabilities in large-scale assessments is controversial, it may be the best way to ensure that their academic progress is monitored and that expectations for their academic performance remain high (Koretz & Barton, 2003).

Koretz and Barton's (2003) review of the literature regarding the inclusion of students with disabilities in large-scale assessments found that very few studies have examined how participation in these assessments has directly affected students with disabilities. They noted concerns that research in this area will be very difficult to conduct due to the inconsistent identification and classification of students with disabilities across states. For example, the proportion of students identified as having a learning disability ranges from 9.1% to almost 16% across states and "such dramatic inconsistencies in identification and classification rates make it difficult to determine how to best assess students with disabilities" (p. 4). Additionally, more research needs to be conducted regarding potential item bias for students with disabilities, as well as the appropriate use of accommodations by students with disabilities. Indeed, research regarding the inclusion of students with disabilities in large-scale assessments is still in its infancy.

### *Old Question, New Methodology*

The results of high-stakes tests for students with disabilities have been used in a variety of studies. For example, students' scores on high-stakes tests have been examined as a way to monitor the ability of students with learning disabilities to achieve grade level proficiency in reading over time (Schulte et al., 2001), to identify appropriate methods for identifying school-level growth in reading (Schulte & Villwock, 2004), to compare methods for assessing trends over time (Ysseldyke & Bielinski, 2002), and to estimate the effectiveness of special education programs (Hanushek et al., 1998; 2002).

Because students with disabilities have only recently been required to participate in large-scale assessments, little research has been conducted regarding their performance on these examinations. A study by Schulte et al. (2001) found that in one North Carolina school district, there was an increase in the proportion of students with learning disabilities (from 39% to 53%) who scored at or above the proficiency level on a large-scale assessment when a value-added model was used to assess growth longitudinally. Despite this growth, a large proportion of these students continued to score below the proficiency cut-off. This study revealed that large-scale assessments can be used to monitor students' growth and proficiency levels over time.

In the same school district, Schulte and Villwock (2004) later used scores on large-scale assessments to determine the percent of students with and without disabilities achieving grade level proficiency each year and the percent of students with and without disabilities who exceeded expected growth each year. These authors concluded that regression-based, value-added analyses can be used effectively to examine the progress of students with disabilities on large-scale assessments. Additionally, state standards for students in general education programs can be applied as goals for students in special education programs. However, they also stated that the entry and exit patterns of special education students, as well as school-level effects for all children create biases that need to be addressed during analyses.

Due to the transitioning of students between special and general education programs, growth over time can be difficult to estimate. Therefore, Ysseldyke and Bielinski (2002) examined the different methods available for examining student achievement growth (e.g., cross-sectional, longitudinal) on a large-scale assessment in the state of Texas. They found that cross-sectional analyses revealed the largest achievement gap between students in special and general education over time. Similar trends were also found when a cohort-dynamic longitudinal method was used (i.e., special or general education status was defined by a student's status in a given grade, such that each student's status could change each year). However, the achievement gap was lessened when a cohort-static longitudinal method was used (i.e., special or general education status was defined by a student's status in the first year of the study and did not change, regardless of whether a student entered into or exited out of special education programs). Because higher achieving students with disabilities were more likely to exit special education programs, over time, the group of special education students demonstrated lower achievement when transitions between programs were taken into account (as was done when both the cross-sectional and cohort-dynamic longitudinal methodologies were used). Thus, the transitioning of students between programs affected group gains over time.

With increased numbers of students with disabilities participating in large-scale assessments, Ysseldyke and Bielinski (2002) stated that "longitudinal data such as that available through some state testing agencies may be very important for assessing the impact of special education services on student achievement" (p. 199). Just as Ysseldyke and Bielinski predicted, recent research by Hanushek et al. (1998, 2002) examining the large-scale achievement scores of students with disabilities who transitioned in or out of special education programs has been influential in providing new data to address longstanding concerns about educational outcomes for students with disabilities.

Hanushek and colleagues' (1998) working paper with the National Bureau of Economic Research was one of the first studies to examine the effectiveness of special education using large-scale assessment data. Based on a desire to understand whether the costs of special education services are justifiable, Hanushek et al. analyzed the reading and math scores of students with disabilities on state-wide assessments in Texas. Their analyses were conducted much differently than traditional special education effectiveness research and their results have played a central role in discussions regarding the rethinking of current special education policy. Through a series of multiple regression analyses, they revealed results that conflicted with the findings of the most frequently cited classic special education studies from the 1980s and 1990s: they demonstrated that special education programs improve the math and reading achievement scores of students with disabilities. In 2002, they published a related study in *The Review of Economics and Statistics* with similar methodology, but with an expanded range of student grade levels. Because Hanushek and colleagues' 1998 study provided a foundation for the current research, their study will now be discussed and critiqued in detail. The next section provides a review of (a) their methodology (e.g., participants, research analyses), (b) the influential aspects of their results, and (c) the need to expand upon this study with a sample that is more representative of the special education population.

*Hanushek et al.'s Use of Large-Scale Assessments to Evaluate Special Education*

*Research Participants*

Hanushek et al. (1998) examined data from the Harvard and University of Texas Dallas (UTD) Texas Schools Project, which included five cohorts of Texas public elementary school students (their data pool included over 200,000 students in more than 3,000 schools in each of the five cohorts). They reported that their sample size varied for each regression analysis done, with fluctuations from 254,294 to 601,526 students participating in any data analysis.

*Research Analyses and Results*

*Normalized TAAS scores.* To assess the effectiveness of special education programs, Hanushek et al. (1998) used students' scores on the Texas Assessment of Academic Skills (TAAS). Between 1993 and 2002, Texas students in the third through eighth grades participated in the TAAS to assess their mastery of grade-level concepts in reading and math (Texas Education Agency, 2003). The availability of state-wide TAAS scores provided an opportunity to assess the effectiveness of special education by predicting students' relative standing on the TAAS from one grade to the next based on their special education status. Because of their emphasis on students with disabilities' gains relative to general education students (rather than on students' ability to achieve proficiency standards), Hanushek et al. began their analyses by first normalizing students' test scores by transforming them to *z*-scores with a mean of 0 and a variance of 1, by year, grade, and subject based on the results of all children taking the test. A similar procedure has been used elsewhere to assess the efficacy of special education programs by examining gains students have made on large-scale assessments (Schulte et al., 2001; Schulte & Villwock, 2004).

*Value-added analyses.* After transforming test scores to *z*-scores, Hanushek et al. (1998) rejected doing a simple between-groups analysis (general education versus special education students) for many of the reasons discussed earlier in this document, such as non-random assignment to groups. Instead, they began their series of analyses with between-groups comparisons of *gain* scores for general versus special education students by grade and subject, with additional analyses looking at gains separately for the three most prevalent disability classifications (learning disabled, speech-language impaired, and emotionally disturbed). They termed these value-added analyses because the dependent measure, gain scores, only reflected gain in the current year, thus controlling for any unchanging variables (such as gender) and previous effects (such as prior achievement) when looking at between-group differences. Specifically, in this initial set of analyses, they used ordinary least squares regression (OLS) to examine the difference in students' annual test score gains as a function of their special education status (i.e., special education versus general education) while controlling for (a) student and family characteristics, (b) school demographic characteristics, and (c) community characteristics.

Results of these analyses revealed no consistent patterns for special education outcomes across grades, subject area, or disabilities. Hanushek et al. (1998) concluded that "these OLS coefficients almost surely confound the true program effect with individual and school influences linked with selection into the special education program" (p. 16).

*Eliminating fixed effects.* As a result of these confounds, Hanushek et al. (1998) rejected their initial results. They concluded that to isolate the true effect of special education, an alternate strategy was needed that compared special education students only to themselves, not to the general education population. To do so, Hanushek et al. focused their primary analyses on a sub-group of students in their dataset who had entered into or exited out of special education programs while in the fourth or fifth grade. Although this new strategy involved focusing on a small subset of all students with disabilities, it allowed for a comparison of the gains for students with disabilities while enrolled in special education programs and while enrolled in general education programs. Although they targeted their analyses toward students who entered or exited special education programs during the 3 years of the study, general education students and students with disabilities who did not transition in or out of special education also were included in the analyses to provide better estimates of the control variables and expected gains in achievement. This approach is termed "differencing" and is common in econometrics (Woolridge, 2000).

*Differenced scores.* For all students, first-level differenced scores were calculated to compare growth in reading and mathematics achievement scores in different grades. The creation of differenced scores involved two steps. First, Hanushek et al. (1998) computed two growth scores: one year of growth was calculated by subtracting the *z*-score for Year 1 from the *z*-score for Year 2 and a second year of growth was calculated by subtracting the *z*-score for Year 2 from the *z*-score for Year 3 (with positive scores indicating gain from one year to the next and negative scores indicating a decrease in scores over time). Then they created differenced scores by subtracting the first year of growth from the second year of growth.

*Estimating special education effectiveness.* With this strategy, Hanushek and colleagues (1998) did not predict gains on TAAS scores based on students' status as a general versus a special education student in their second set of analyses. Instead, they predicted differences in students' TAAS score gains with a transition between special and general education placement. Although students who remained in general or special education programs across the 3 years were included in the analyses for comparison, the groups of most interest were students with disabilities transitioning into (entering) special education from general education programs or transitioning out of (exiting) special education to general education programs. By examining gains made in the special education versus general education environment for students with disabilities who entered or exited special education programs, these authors eliminated the biases associated with comparing special to general education students' scores.

This second set of OLS multiple regression analyses examined the difference in students' test score gains from year to year as a function of (a) their special education status (i.e., entered into special education programs, exited out of special education programs, or were always in general education or special education programs), (b) school characteristics (the percent of students who receive a free or reduced price lunch, the percent of students who are Black, and the percent of students who are Hispanic), and (c) random error. Because each student served as his or her own control, the comparison groups' demographic characteristics were stable, and it was not necessary for the authors to continue to include family or community characteristics in the regression equation.

*Results of the differenced regression analyses.* The new regression analyses were run in two different ways. First, students who entered or exited special education programs were considered together as a group and their differenced scores were compared to those for students who consistently remained in special or general education programs. Next, the effects of entering special education programs and exiting special education programs were considered separately by deleting one group or the other from the analyses. Hanushek et al. (1998) were primarily interested in the effects for students who entered into special education programs and were less interested in the effects of both special education transitions combined. They expected that after a student exited a special education program, residual effects of the special education instruction would likely continue to impact students' performance even after they were dismissed from these services and enrolled in general education programs. Therefore, the difference between gains made during special versus general education programs by students who entered special education programs would be most indicative of the effectiveness of special education programs. Table 1 provides a summary of the special versus general education placement of students across the 3 years of the study.

Table 1

*Placement of Students by Group and Year*

| | Placement | Relevant Comparison for Special Education |
|---|---|---|
| *Entered special education* | | |
| Year 1 | General Education | |
| Year 2 | General Education | |
| Year 3 | Special Education | (**Y3-Y2**) – (Y2-Y1) |
| *Exited special education* | | |
| Year 1 | Special Education | |
| Year 2 | Special Education | |
| Year 3 | General Education | (Y3-Y2) – (**Y2-Y1**) |
| *Always special education* | | |
| Year 1 | Special Education | |
| Year 2 | Special Education | |
| Year 3 | Special Education | NA |
| *Always general education* | | |
| Year 1 | General Education | |
| Year 2 | General Education | |
| Year 3 | General Education | NA |

*Note.* Special education gain scores are printed in **bold** lettering.

After completing these three analyses (both transitions combined, special education entry only, special education exit only) for all disability categories combined, the same analyses were run separately for students in each of the three largest disability categories. Hanushek et al. (1998) predicted “special education to have its largest achievement impact on learning disabled students and a much smaller impact on students classified as speech impaired” (p. 15) arguing that speech services were not as targeted toward academics, particularly in the area of math.

Overall, the results revealed that special education increased students' TAAS scores in both math and reading. When achievement gains while in special education were considered separately for students who had entered versus exited special education, all regression coefficients for entrants into special education were statistically significant, and all regression coefficients for exiters from special education were not. This pattern was true for both reading and math scores. Thus, based on the results for students entering special education programs, Hanushek et al. (1998) concluded that special education programs are associated with increases in students' academic achievement. Interestingly, the effects for math for entrants were found to be almost twice as large as the effects for reading. In fact, when Hanushek and colleagues published the findings of their second, related study in *The Review of Economics and Statistics* in 2002, they only reported the analyses conducted with math scores.

In addition to the overall finding supporting the hypothesis that special education programs demonstrate effectiveness for students with disabilities, the authors revealed a few other interesting results. First, when limiting the sample to include students who did not change schools, the regression coefficient remained similar, suggesting that transferring between schools did not seem to bias the results. Second, the results for each of the three main disability types were similar to the results found across disability types. As predicted, the results for students with learning disabilities (the largest of the disability categories) were most similar to the entire group of special education students. Also of interest was the finding that the students with learning disabilities (and to a lesser extent, students with emotional disturbance) made much greater math gains that did students with speech impairments. Hanushek and colleagues (1998) believe this finding provided additional support for the effectiveness of special education for students with learning disabilities.

*Examining potential biases.* Due to concerns about interpretation of scores when students transition in and out of special education programs, Hanushek et al. (1998) then conducted a series of follow-up analyses to rule out potential sources of bias in their analyses. Through the use of both logical arguments and additional multiple regression analyses, these authors showed that biases in terms of (a) students entering special education programs based on only a single year of temporarily low achievement scores (i.e., students who had a bad testing day would likely experience greater gains the following year than would students who consistently scored lower over the course of multiple years); (b) schools manipulating which students take the test; or (c) differences in school quality across grades were unlikely to have affected the results. In sum, these analyses demonstrated that special education placement is associated with improvements in the math and reading achievement scores of students with disabilities.

*Why This Research Was Influential*

The analyses by Hanushek and colleagues (1998) greatly improved upon previous special education effectiveness studies for a variety of reasons, including using large-scale assessment data as an outcome variable, eliminating biases associated with the use of non-equivalent control groups, having a very large sample size, and conducting a number of follow-up analyses to assess for potential biases. Each of these influential pieces of their study will be discussed in more detail.

First, Hanushek and colleagues (1998) considered it important to use students' scores on a statewide, high-stakes assessment of achievement as their outcome measure. Legislative mandates such as the 1997 reauthorization of the IDEA and the NCLB Act require that schools demonstrate adequate progress in improving the achievement for children in both general and special education programs by documenting their scores on large-scale academic assessments. Thus, Hanushek et al.'s use of differenced scores on the TAAS (and the current study's use of first-level differenced scores on the large-scale assessment used in North Carolina) as an outcome variable is logical because of the emphasis current legislation places on students' scores on large-scale assessments.

Next, Hanushek and colleagues (1998) recognized that the inherent differences between students enrolled in special and general education programs create unavoidable biases when special education effectiveness studies utilize control groups involving general education students. Therefore, their "control group" consisted of the special education students during a year in which they were not classified as needing special education. Their primary multiple regression formula involved examining students' test score gain as a function of their special education status, school characteristics, and random error. The coefficient for special education status did not indicate whether students were placed within a general or special education program; rather, it represented whether a student entered or exited a special education program during the study. Hanushek et al. did not randomly assign participants to groups; however, by focusing the special education outcomes for students who transitioned in or out of special education programs, they reduced (but did not eliminate) the effects of non-random assignment of students to treatment and comparison groups.

Additionally, Hanushek and colleagues' (1998) research was influential because of the large sample of participants used. Their use of data from five cohorts of Texas elementary school students allowed them to maintain a large sample size.

Finally, Hanushek and colleagues (1998) conducted a variety of follow-up analyses to provide additional support for their finding that special education placement was associated with increased TAAS scores for students with disabilities. These follow-up analyses revealed no evidence of bias in terms of students being temporarily shocked by a single year of low achievement scores, teachers manipulating who takes the TAAS tests, or the school quality changing across grades and years. Although they admitted that other sources of bias may be possible, these analyses provided additional support for their results.

This important study has influenced a variety of other researchers. According to a recent cited reference search on the Social Sciences Citation Index, at least a dozen other research articles have cited this 1998 working paper. Additionally, in the 7 years since Hanushek and colleagues published their related study in *The Review of Economics and Statistics* in 2002, it has already been cited an additional three times. Thus, this important research has influenced a variety of researchers interested in studying special education policy.

*The Need to Replicate and Extend This Research*

Despite its impact on the field of special education, Hanushek and colleagues' (1998) research has yet to be replicated. Although these findings are quite powerful, a replication of the analyses with data from another state would greatly increase the external validity of the findings. At the time of their study, only 30% of students with disabilities took the TAAS. Thus, in 2002, the authors themselves wondered about the generalizability of their data to the special education programs in other states by arguing that, "some question arises whether the results obtained from the tested population are generalizable to all students who receive special education" (p. 588). The current study re-examines the effectiveness of special education programs in a state in which a much larger proportion of students with disabilities participate in large-scale assessments. Furthermore, in 2003, *Princeton Review* ranked North Carolina's testing and accountability program fourth in the nation (just behind the state of Texas), making its data ideal for use in a replication of this kind.

Additionally, Hanushek and colleagues (1998) did not adequately account for the possibility that the scores for students in special education programs may regress toward the mean. All observed scores contain measurement error that inflates or deflates students' scores away from their true scores. Students with extreme scores below the mean are more likely to have observed scores that are deflated due to measurement error. Upon retesting, these students' scores are likely to include measurement error that will inflate their scores toward the mean. Thus, students' low pretest scores will increase at posttest simply due to measurement error, rather than the effects of a treatment (e.g., special education; Leary, 2004). In their 1998 working paper, Hanushek et al. excluded the top and bottom 1% of their participants' gain scores from their analyses because of "concerns about measurement error." Using the same reasoning, they excluded the bottom 1% of participants' scores because these students scored lower than one would when randomly guessing. However, regression to the mean can be accounted for within the regression analysis itself, rather than by simply excluding participants with extreme scores. In fact, North Carolina adjusts for regression to the mean in their state-wide formula for calculating students' growth on large-scale assessments [Public Schools of North Carolina (PSNC), 2004c]. Because the participants include students with disabilities who are more likely to receive extreme scores, the current study compared analyses conducted in the same manner as was done by Hanushek et al. to analyses that also account for potential regression to the mean.

Hanushek et al. (1998) also did not address the fact that students receiving special education services are frequently given accommodations when taking large-scale assessments. It is possible that these students scored higher on the TAAS when they were enrolled in a special education program because their special education status permitted them to receive accommodations during the test that they did not receive in the year in which they were not classified as a special education student. Therefore, in the current study, accommodation effects are examined as well.

Finally, Hanushek and colleagues (1998) disregarded the fact that their results revealed less significant outcomes for students on the reading portion of the TAAS than on the math portion of the TAAS. Lyon et al. (2001) reported that 80% of students with learning disabilities experience difficulties in reading as their primary academic difficulty. Additionally, a significant amount of educational research has revealed more information about the process of reading than about any other academic domain in which a student can be identified as having a learning disability. Because the majority of students with specific learning disabilities are receiving services for reading difficulties, the less strong results found by Hanushek et al. in the area of reading should not be ignored (or excluded completely, as was done in their 2002 publication).

In sum, Hanushek and colleagues (1998) found that entry into special education programs resulted in an overall increase in large-scale achievement scores for Texas students with disabilities. The current study replicates and expands upon this study by examining the scores of North Carolina students with disabilities on a state-wide test of achievement.

# CHAPTER THREE

## Research Aims

### *Statement of the Problem*

As is evident from the literature review in the previous chapter, questions regarding the education of students with disabilities have changed from questions of access to questions of outcomes. Over time, federal mandates such as the original IDEA, the Goals 2000: Educate America Act, the 1997 reauthorization of IDEA, and the NCLB Act have increasingly provided students with disabilities access to the general education curriculum and increased the accountability of schools and states to provide these students with an appropriate education. As federal legislation changed, these students' access to an education was deemphasized and the need to examine academic outcomes for these students became a more pressing concern.

The effectiveness of special education programs has been an ongoing question in the educational community. Over time, this question has been addressed by examining outcomes for students in special education programs, frequently in comparison to outcomes for students in general education programs. Literature reviews and meta-analyses have revealed mixed results regarding the effectiveness of special education programs for students with varying severity of disabilities. Overall, however, outcomes for students in special education programs have historically not been encouraging.

Recently, students with disabilities have been increasingly included in large-scale assessments of academic achievement as a way to hold schools, school districts, and states accountable for their education. This new method for assessing outcomes for students with disabilities has provided researchers with an alternate means for examining the effectiveness of special education programs. Research by Hanushek et al. (1998, 2002) examined the effectiveness of special education programs by examining the differenced scores of students who transitioned in or out of special education programs. This unique methodology allowed for the evaluation of special education practices across the entire state of Texas.

The current study replicates and extends the special education effectiveness study by Hanushek et al. (1998) using data collected from elementary school students in North Carolina. Similar to the students in Texas, each North Carolina student in grades three through eight participates in End-of-Grade (EOG) Tests of Reading Comprehension and Mathematics (PSNC, 1998). Since 1997, the North Carolina Department of Public Instruction has kept a data file including the EOG Test scores for students at each North Carolina public school. Recently, the North Carolina Education Research Data Center (NCERDC) created one file for each grade by combining the separate school files for both tests. These data files are currently housed within the Center for Child and Family Policy at Duke University. The current study involves the use of these files to run multiple regression analyses that test the external validity of Hanushek and colleagues' findings. Because a much higher proportion of students with disabilities take the North Carolina EOG Tests each year, this study assesses the generalizability of the results found in Texas using a more representative sample of the special education population. The results of these analyses have the potential to greatly influence the field of education by either supporting or challenging the claims made by Hanushek and colleagues that special education significantly improves the achievement scores of students with disabilities.

## *Hypotheses*

### *Hypothesis 1*

(a) Receipt of special education services will be a positive predictor of academic gain in the area of reading comprehension.

(b) Receipt of special education services will be a positive predictor of academic gain in the area of mathematics.

Hypotheses 1a and 1b will be tested using a multiple regression equation that is very similar to Hanushek and colleagues' (1998) second formula, which eliminated fixed individual effects by predicting students' achievement gain in one year relative to another year (i.e., differenced scores) from their special education transition status (i.e., whether they were exiting special education, entering special education programs, or did not transition). Three regression equations will be used for testing both 1a and 1b, with the first analysis looking at gains for all transitioning (i.e., entering and exiting) students followed by separate analyses for entering students alone and exiting students alone. It is hypothesized that the results will be similar to the results found by Hanushek et al., providing support for the effectiveness of special education.

Although not formally part of Hypothesis 1, follow-up analyses will examine results separately for students from the four largest disability categories represented in the dataset [specific learning disabled (SLD), speech-language impaired (SLI), behaviorally-emotionally disabled (BED), and other health impaired (OHI)]. Hanushek et al. (1998)

looked at the first three disability subgroups separately, but did not examine results for the OHI category. It will be examined separately in the present study because it is one of the most prevalent disability categories in North Carolina.

*Hypothesis 2*

(a) The advantages of special education over general education placement for students in the area of reading comprehension will continue to exist after controlling for regression to the mean.

(b) The advantages of special education over general education placement for students in the area of mathematics will continue to exist after controlling for regression to the mean.

Both parts of the second hypothesis will be tested with two regression equations so that gains for students who entered and exited special education programs can be considered separately. It is hypothesized that even after controlling for regression to the mean, support for the effectiveness of special education will continue to be found. Again, although not part of the formal hypothesis, separate analyses will also be conducted for both the entire group of students and students from the four previously specified disability categories.

*Hypothesis 3*

(a) The advantages of special education over general education placement in the area of reading comprehension will continue to exist after controlling for the effects of using accommodations by students in special education programs.

(b) The advantages of special education over general education placement in the area of mathematics will continue to exist after controlling for the effects of using accommodations by students in special education programs.

Hanushek et al. (1998) did not consider the effects of receiving accommodations on large-scale assessments on the gain scores of students with disabilities. Thus, the current study will examine the effectiveness of special education even after this confound is controlled. It is expected that after controlling for the use of accommodations by special education students, the effectiveness of special education will continue to be apparent. Again, although not part of the formal hypothesis, separate analyses will also be conducted for both the entire group of students and students from the four previously specified disability categories.

# CHAPTER FOUR

## Method

### *Study Overview*

Extant data from the North Carolina Department of Public Instruction’s reading and mathematics large-scale assessment program were used in the current study to estimate the effectiveness of special education services. Within North Carolina’s testing program, students in public and charter schools are tested annually. These data are available for secondary analyses through the North Carolina Education Research Data Center with pupil and teacher identifiers removed. Files containing achievement test scores and other data used in the present study were accessed following study approval by the Center and North Carolina State University’s Institutional Review Board.

### *Participants*

Large scale achievement test scores from five cohorts (1997-2003) of North Carolina elementary school students were used in the present study. For the primary analyses, fifth grade test scores of students who entered or exited special education programs between the fourth and fifth grade or remained in special or general education programs from the third through the fifth grades were included. Only students from this group who: (a) had taken the North Carolina End-of-Grade Tests in Reading Comprehension and/or Mathematics (EOG-Reading and EOG-Math) while in fifth grade, (b) had also taken the fourth grade EOG-Reading and/or EOG-Math Tests the year prior to entering fifth grade, (c) had also taken the third grade EOG-Reading and/or EOG-Math Tests the year prior to entering fourth grade, and (d) had school-level demographic information that was available (or could be approximated), were included in the primary analyses.

The final total sample size was 396,828 students who had at least one set of complete third, fourth, and fifth grade EOG Test scores available in the areas of reading and/or mathematics. Of the 396,828 participating students, 6,569 entered into special education programs during the fifth grade, 5,043 exited out of special education programs during the fifth grade, 29,878 remained in special education programs throughout the third through fifth grades, and 355,338 remained in general education programs throughout the third through fifth grades. Demographic characteristics of these students are presented in Table 2. The preliminary analyses section of Chapter Five will provide detailed information on the decision rules for including and excluding students in the study sample and their impact on sample size.

Table 2

*Demographic Characteristics of the Study Sample*

| | Entered special education | | Exited special education | | Always special education | | Always general education | |
|---|---|---|---|---|---|---|---|---|
| | *N* | % | *N* | % | *N* | % | *N* | % |
| *Gender* | | | | | | | | |
| Female | 2,363 | 36.0% | 1,784 | 35.4% | 8,911 | 29.8% | 188,536 | 53.1% |
| Male | 4,206 | 64.0% | 3,259 | 64.6% | 20,963 | 70.2% | 166,792 | 46.9% |
| *Ethnicity* | | | | | | | | |
| White | 3,871 | 58.9% | 3,338 | 66.2% | 19,230 | 64.4% | 224,786 | 63.3% |
| Black | 2,260 | 34.4% | 1,426 | 28.3% | 9,134 | 30.6% | 104,062 | 29.3% |
| Hispanic | 186 | 2.8% | 111 | 2.2% | 620 | 2.1% | 10,553 | 3.0% |
| *Disability* | | | | | | | | |
| SLD | 3,813 | 58.0% | 2,220 | 44.0% | 19,349 | 64.8% | 0 | 0.0% |
| SLI | 736 | 11.2% | 2,304 | 45.7% | 1,719 | 5.8% | 0 | 0.0% |
| BED | 439 | 6.7% | 133 | 2.6% | 2,116 | 7.1% | 0 | 0.0% |
| OHI | 1,057 | 16.2% | 186 | 3.7% | 4,089 | 13.7% | 0 | 0.0% |
| Other | 524 | 7.9% | 200 | 4.0% | 2,605 | 8.7% | 0 | 0.0% |
| *Lunch Price* | | | | | | | | |
| Free | 2,717 | 41.4% | 1,860 | 36.9% | 12,326 | 41.2% | 106,197 | 29.9% |
| Reduced | 461 | 7.0% | 372 | 7.4% | 2,198 | 7.4% | 23,966 | 6.7% |
| *Cohort* | | | | | | | | |
| 1998-99 | 1,289 | 19.6% | 986 | 19.6% | 5,718 | 19.1% | 67,618 | 19.0% |
| 1999-00 | 1,421 | 21.6% | 975 | 19.3% | 5,901 | 19.8% | 70,321 | 19.8% |
| 2000-01 | 1,413 | 21.5% | 977 | 19.4% | 5,802 | 19.4% | 71,655 | 20.2% |
| 2001-02 | 1,219 | 18.7% | 1,002 | 19.9% | 5,951 | 19.9% | 72,157 | 20.3% |
| 2002-03 | 1,289 | 19.6% | 1,103 | 21.9% | 6,506 | 21.8% | 73,587 | 20.7% |
| *Total* | 6,569 | | 5,043 | | 29,878 | | 355,338 | |

*Measures*

The outcome variables for all analyses were students' first-level differenced gain scores on the EOG-Reading and EOG-Math Tests. The EOG-Reading and Math Tests are currently used for accountability purposes across the state of North Carolina. As part of the North Carolina testing program, these two tests are used to measure how well students meet academic performance requirements in reading comprehension and mathematics. Scores are reported in percentiles, scale scores, and achievement levels. Percentiles are based on the population of students who took the test in the norming year. Raw scores (total number of items correct) are translated into developmental scale scores, which allow students' growth across grades to be reflected on a single scale. Developmental scale scores also are used to place students into one of four achievement levels, indicating a student's mastery of grade-level knowledge (PSNC, 2004b; 2006). Grade-level proficiency is defined as scoring at or above Achievement Level III on both the EOG-Reading and the EOG-Math Tests (PSNC, 2004b; 2006).

*EOG Test of Reading Comprehension.* The North Carolina EOG Test of Reading Comprehension assesses students' ability to read and comprehend written material that is appropriate for students performing at a particular grade level. The test includes ten reading passages that are each followed by three to eight multiple choice comprehension questions. The third through fifth grade tests include 50 questions to be answered in 115 minutes. Each test is comprised of four literary passages, four content-based passages, and two consumer/human interest passages. The test measures four concepts: (a) cognition (e.g., understanding the purpose and organization of a reading passage), (b) interpretation (e.g., clarifying, explaining, expanding ideas found in a reading passage), (c) critical stance (e.g., comparing and contrasting), and (d) connections (e.g., connecting information found in a reading passage with other information and experiences; PSNC, 1998; 2004b).

Internal consistency reliability estimates (coefficient alphas) for the EOG-Reading Tests range from .92 to .94 for grades three through eight. These estimates are slightly lower for the third grade pretest (.82) and the tenth grade test (.88). Additionally, the standard error of measurement for all students taking the test is two to six points; the standard error of measurement for 95% of the students taking the test (who score within two standard deviations of the mean) is two to three points. Thus, the EOG-Reading Tests demonstrate adequate reliability (PSNC, 2004b).

Evidence of both content and criterion-related validity exists for the EOG-Reading Tests. Content validity was demonstrated through analyses of the judged match between item content and the four basic constructs involved in higher-order thinking that are emphasized in the curriculum: cognition, interpretation, critical stance, and connections. For the third grade test, approximately 37% of the items measure cognition, 37% of the items measure interpretation, 19% of the items measure critical stance, and 7% of the items measure connections. For the fourth grade test, approximately 40% of the items measure cognition, 38% of the items measure interpretation, 18% of the items measure critical stance, and 4% of the items measure connections. For the fifth grade test, 35% of the items measure cognition, 40% of the items measure interpretation, 20% of the items measure critical stance, and 5% of the items measure connections. Criterion-related validity was demonstrated through moderate to strong Pearson correlation coefficients with associated variables (e.g., assigned achievement level by expected grade, teacher judgment of achievement by assigned achievement level, expected grade by scaled score, and teacher judgment of achievement by scaled score; PSNC, 2004b).

*EOG Test of Mathematics.* The North Carolina EOG Test of Mathematics assesses students' ability to apply mathematical principles, solve mathematical problems, and explain mathematical processes using problems in context. The test is comprised of two sections: Calculator Inactive and Calculator Active. The Calculator Inactive section assesses students' ability to perform mathematical computations without a calculator. This section includes 28 items to be completed in 60 minutes for third through fifth grade students. The Calculator Active section of the test allows for the use of calculators, rulers, and protractors and includes 54 items to be completed in 135 minutes for third through fifth grade students. Thus, the entire EOG-Math Test includes 82 items to be completed in 195 minutes. The test measures four concepts: (a) number sense, numeration, and numerical operations; (b) spatial sense, measurement, and geometry; (c) patterns, relationships, and functions; and (d) data, probability, and statistics (PSNC, 1998; 2004a).

Internal consistency reliability estimates (coefficient alphas) for the EOG-Math Tests range from .94 to .96 for grades three through eight and grade ten. The estimate is slightly lower for the third grade pretest (.82). These coefficient alphas remained the same when examined separately by gender, ethnicity, disability status, or Limited English Proficiency status. Similar to the EOG-Reading Tests, the standard error of measurement for all students taking the test is two to six points; the standard error of measurement for 95% of the students taking the test (who score within two standard deviations of the mean) is two to three points. Thus, the EOG-Math Tests demonstrate adequate reliability (PSNC, 2006).

Evidence of both content and criterion-related validity exists for the EOG-Math Tests. Content validity was demonstrated through analyses of how well the items measured the four basic constructs involved in higher-order thinking that are emphasized in the curriculum: number sense, numeration, and numerical operations; spatial sense, measurement, and geometry; patterns, relationships, and functions; and data, probability, and statistics. North Carolina teachers also evaluated the math test items on a number of additional criteria. Items included on the tests conformed to the following criteria to a “high” or “superior” degree according to teacher ratings: (a) tapped grade-level curriculum objectives; (b) reflected the grade-level curriculum taught in their individual school; (c) were clearly and concisely written; (d) were not biased against students of different races, genders, socioeconomic statuses, or geographic locations; and (e) had only one answer. Criterion-related validity was demonstrated through moderate to strong Pearson correlation coefficients with associated variables (e.g., assigned achievement level by expected grade, teacher judgment of achievement by assigned achievement level, teacher judgment of achievement by expected grade, teacher judgment of achievement by math scale score, and expected grade by math scale score; PSNC, 2006).

*Procedure*

*Test administration.* North Carolina EOG Tests are administered to students in May of each year. Between 1997 and 2003, the tests typically were administered by general education teachers and proctors in the students’ general education classrooms.

The EOG-Reading and EOG-Math Tests were given in the mornings on separate days during the course of a week (Schulte et al., 2001).

State testing guidelines require students who are pursuing the state's standard course of study be included in the large scale testing, regardless of special education status. However, there are procedures to exempt students from testing 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 the testing (PSNC, 2007). Participation rates in the testing for special education students varied by year. Across the 5 years included in the study, the overall participation rate for the fourth and fifth grades ranged from 73% to 84% on the EOG-Reading Tests, and 78% to 86% on the EOG-Math Tests.

In addition to exemption from the reading and/or the mathematics testing, students could receive a variety of accommodations during the EOG Tests, if use of the accommodations had been specified in their IEP, Section 504 Plan, or Limited English Proficiency Plan (PSNC 2004b; 2006). Available accommodations included having extended time, taking the test in a separate setting, having a translator, using an abacus, using a test with large print, and marking in the test booklet. Within the overall sample of students with disabilities, the most commonly provided accommodations were: receiving extended time on the test, taking the test in a separate room, having the test read aloud, marking in the test booklet, and taking the test in multiple sessions. Although accommodations are generally provided only to students receiving special education services, some students participating in general education programs also received these accommodations. Table 3 presents the percent of students in the each subgroup of the sample (those entering special education programs during the fifth grade, those exiting special education programs during the fifth grade, those who were in special education programs throughout the study, and those who were in general education programs throughout the study) who received each of these accommodations in the third, fourth, and fifth grades.

Table 3

*The Percent of Students Who Used the Five Most Commonly Provided Accommodations, by Group and Year*

| | Test Accommodation | | | | |
|---|---|---|---|---|---|
| Students | Extended Time | Separate Room | Read Aloud | Marking in Booklet | Multiple Sessions |
| *Entered special education* | | | | | |
| Grade 3 | 6.23% | 6.42% | 4.40% | 2.92% | 1.51% |
| Grade 4 | 11.25% | 10.41% | 6.21% | 5.21% | 2.94% |
| Grade 5 | 67.28% | 63.27% | 42.70% | 36.92% | 21.36% |
| *Exited special education* | | | | | |
| Grade 3 | 38.33% | 38.61% | 29.66% | 22.67% | 9.00% |
| Grade 4 | 38.75% | 35.26% | 25.12% | 20.21% | 8.96% |
| Grade 5 | 12.77% | 10.71% | 6.70% | 6.07% | 3.07% |
| *Always in special education* | | | | | |
| Grade 3 | 70.41% | 73.22% | 62.62% | 45.00% | 19.04% |
| Grade 4 | 75.36% | 73.81% | 61.87% | 46.42% | 22.65% |
| Grade 5 | 78.55% | 73.67% | 60.09% | 47.24% | 24.73% |
| *Always in general education* | | | | | |
| Grade 3 | 1.71% | 1.52% | 1.19% | 0.73% | 0.36% |
| Grade 4 | 1.96% | 1.67% | 1.10% | 0.79% | 0.37% |
| Grade 5 | 2.19% | 1.80% | 0.99% | 0.93% | 0.60% |

Although some students did receive testing accommodations when not classified as having a disability, 97 – 98% of students who were always in general education programs (and were never identified as having a disability during the years examined) did not receive any accommodations on the tests each year. Table 4 summarizes students' use of accommodations on the EOG Tests by participants in the present study.

Table 4

*Students' Use of Accommodations by Group*

| Variable | Mean number of accommodations | Standard deviation | Range of accommodations |
|---|---|---|---|
| *Entered special education* | | | |
| Grade 3 | 0.20 | 0.79 | 0 – 6 |
| Grade 4 | 0.38 | 1.08 | 0 – 7 |
| Grade 5 | 2.40 | 1.76 | 0 – 8 |
| *Exited special education* | | | |
| Grade 3 | 1.31 | 1.65 | 0 – 6 |
| Grade 4 | 1.35 | 1.72 | 0 – 7 |
| Grade 5 | 0.41 | 1.10 | 0 – 6 |
| *Special education* | | | |
| Grade 3 | 2.62 | 1.60 | 0 – 9 |
| Grade 4 | 2.97 | 1.66 | 0 – 9 |
| Grade 5 | 2.98 | 1.63 | 0 – 11 |
| *General education* | | | |
| Grade 3 | 0.05 | 0.41 | 0 – 10 |
| Grade 4 | 0.06 | 0.47 | 0 – 10 |
| Grade 5 | 0.07 | 0.48 | 0 – 10 |

*Determination of special education status.* The identification of students as disabled and in need of special education services occurred at the school level, based on procedures specified in the state's *Policies Governing Services for Children with Disabilities* (PSNC, 2007). Students were classified into 1 of 12 categories in terms of their exceptionality: (a) not identified as an exceptional student, (b) academically/intellectually gifted, (c) behaviorally-emotionally disabled, (d) hearing impaired, (e) educable mentally handicapped, (f) specific learning disabled, (g) speech-language impaired, (h) visually impaired, (i) other health impaired, (j) orthopedically impaired, (k) traumatic brain injury, and (l) other exceptional classifications. For several of the primary analyses in the present study, students in the first two categories were grouped together as students without disabilities (i.e., general education students) and students in the remaining ten categories were grouped together as students with disabilities (i.e., students participating in special education programs).

*School characteristics.* Similar to the research by Hanushek et al. (1998), three school-level variables were included in analyses for the present study: the percent of students receiving a free or reduced price lunch, the percent of students who are Black, and the percent of students who are Hispanic. This information regarding school characteristics was imported for each cohort from a separate school-level database at the NCERDC. In a small number of cases, information from the National School Lunch Program was missing from school-level files at the NCERDC, and was estimated using student-level files cumulated by school. More information on this data estimation procedure is provided in the preliminary data analyses section of Chapter Five.

# CHAPTER FIVE

## Results

This chapter summarizes the data analyses conducted and the results of these analyses. First, the preliminary data preparation steps and analyses are described and the resulting descriptive statistics are provided. Preliminary data preparation steps included merging student data across the five cohorts, merging school data with the student-level data, transforming students' test scores into normalized *z*-scores, creating differenced scores to estimate the effectiveness of special education programs, and creating differenced residualized gain scores used to correct estimates of achievement gains for regression toward the mean. Second, the data analysis procedures and results related to the three hypotheses presented in Chapter Three are presented.

### *Descriptive and Preliminary Analyses*

As noted in the Method section, the present study made use of extant achievement test data for students in the state of North Carolina who had available test scores from the third, fourth, and fifth grades. However, scores from the entire population of third, fourth, and fifth grade students tested during the target years were also used in the preliminary analyses. Use of all students' scores was necessary to create grade-based *z*-scores, as well as the residual gain scores used in some of the analyses. A complete description of the procedures used to create *z*-scores, gain scores based on *z*-scores, residual gain scores, and differenced scores; descriptive statistics for test scores across years; extent of missing test data; and decisions to exclude test scores is provided in this section. This information is provided in order to permit independent evaluation of the generalizability of the results to broader populations of students with disabilities.

*Merging Data Across Cohorts*

In the NCERDC database, student-level data are recorded by year. Therefore, the first step in the preliminary analyses was to merge students' third, fourth, and fifth grade data to create files for each of the five cohorts that contained data reflecting 3 years of test scores (i.e., 2 years of growth) for each student in the cohort. Then, these five cohorts were merged together to create one large database of student information from 1997 to 2003. During the 5-year time period examined, 515,095 unique student records for fifth graders were available in the NCERDC database for creating yearly grade-based *z*-scores.

One problem that emerged in merging cohort data was that when a student was retained in the third, fourth, or fifth grade, he or she appeared in the same grade for 2 subsequent years. When this occurred, only adjacent third, fourth, and fifth grade scores for the student were kept in the present database. That is, if a student was promoted from third grade to fourth grade to fifth grade and then repeated the fifth grade, only the third grade scores, fourth grade scores, and the first year of fifth grade scores were used in the analyses. Similarly, if a student was retained in the third grade before being promoted to the fourth and fifth grades, only the second year of third grade scores, the fourth grade scores, and the fifth grade scores were used in the analyses. However, if a student repeated the fourth grade before being promoted to fifth grade, the student was not retained in the database because these students represented a small number of the available records ($n$ = 11,115) and there was no accurate way to calculate residual gain scores with two fourth grade scores. There were an additional 81,269 students who were missing data from both EOG Tests for one or more years of the current study. After excluding these 92,384 students, 422,711 student records were found to have complete third, fourth, and fifth grade scores available in the area of reading and/or mathematics.

*Normalizing Raw Test Scores and Calculating Gain Scores*

To complete the planned analyses, it was necessary to create normalized *z*-scores for students based on the mean and standard deviation for each year and grade for all students tested, regardless of whether the students were in special education programs. These scores were then used in the calculation of gain scores and residual gain scores.

First, the means and standard deviations of the EOG-Reading and EOG-Math Test scores were calculated in the original NCERDC student files by year and grade. Next, *z*-scores were created by subtracting the mean EOG Test score for the appropriate year and grade from students' individual scores, and dividing by the overall standard deviation for the appropriate year and grade. Fourth grade (Year 1) growth scores were calculated by subtracting students' third grade pre-test test *z*-scores from their fourth grade post-test *z*-scores; fifth grade (Year 2) gain scores were calculated by subtracting students' fourth grade pre-test *z*-scores from their fifth grade post-test *z*-scores.

*Creating Residualized Gain Scores*

Regression to the mean was controlled for through the use of standard residualized gain scores. To create the residual gain scores for the fourth grade, the correlation between students' third grade and fourth grade test scores was determined for all students tested, by cohort; to create the residual gain scores for the fifth grade, the correlation between students' fourth grade and fifth grade test scores was determined for all students tested, by cohort. The correlations used to create these residual gain scores are presented in Table 5. Students' predicted fourth grade test scores were derived by multiplying their third grade test *z*-score by the appropriate correlation coefficient; students' predicted fifth grade test scores were derived by multiplying their fourth grade test *z*-score by the appropriate correlation coefficient. The difference between students' predicted and actual test scores constituted their residual gain scores for each year.

Table 5

*Correlations Used to Compute Residual Gain Scores*

| | Correlation coefficients | |
|---|---|---|
| Cohorts | Reading | Math |
| Fourth grade (Year 1) gain | | |
| Cohort 1: 1997-1998 | .83 | .82 |
| Cohort 2: 1998-1999 | .82 | .82 |
| Cohort 3: 1999-2000 | .81 | .82 |
| Cohort 4: 2000-2001 | .81 | .81 |
| Cohort 5: 2001-2002 | .80 | .83 |
| Fifth grade (Year 2) gain | | |
| Cohort 1: 1998-1999 | .84 | .85 |
| Cohort 2: 1999-2000 | .83 | .84 |
| Cohort 3: 2000-2001 | .82 | .84 |
| Cohort 4: 2001-2002 | .81 | .85 |
| Cohort 5: 2002-2003 | .81 | .85 |

Tables 6, 7, 8, and 9 report the mean *z*-scores, *z*-score gain scores, and residual gain scores on the EOG-Reading and EOG-Math Tests for students entering special education, students exiting special education, students consistently participating in special education programs, and students consistently participating in general education programs. The mean *z*-scores, *z*-score gain scores, and residual gain scores are reported by cohort, as well as across cohorts, for each group of students. Tables 10 and 11 report the mean *z*-scores, *z*-score gain scores, and residual gain scores on the EOG-Reading and EOG-Math Tests for the sample of students entering into or exiting out of special education programs who were labeled as having a specific learning disability, speech-language impairment, behavioral-emotional disability, or other health impairment. The percent of students for whom test data were available are presented in Table 12; participation rates are reported by cohort, as well as across cohorts, for each group of students.

Table 6

*z-Score Means and Standard Deviations for Students Entering Into Special Education by Cohort and Across All Cohorts*

| | Cohort year | | | | | |
|---|---|---|---|---|---|---|
| | 97-99 | 98-00 | 99-01 | 00-02 | 01-03 | All |
| *Reading comprehension* | | | | | | |
| 3rd grade *z*-score | -0.72 | -0.70 | -0.65 | -0.64 | -0.61 | -0.67 |
| | (0.89) | (0.91) | (0.94) | (0.96) | (1.02) | (0.94) |
| 4th grade *z*-score | -0.86 | -0.80 | -0.77 | -0.76 | -0.76 | -0.79 |
| | (0.91) | (0.93) | (0.97) | (0.98) | (1.07) | (0.97) |
| 5th grade *z*-score | -0.81 | -0.71 | -0.66 | -0.61 | -0.57 | -0.67 |
| | (0.96) | (0.97) | (1.02) | (1.00) | (1.02) | (1.00) |
| 4th grade *z*-score gain | -0.13 | -0.10 | -0.12 | -0.12 | -0.15 | -0.13 |
| | (0.67) | (0.67) | (0.68) | (0.69) | (0.66) | (0.67) |
| 5th grade *z*-score gain | 0.03 | 0.08 | 0.09 | 0.10 | 0.18 | 0.10 |
| | (0.68) | (0.71) | (0.70) | (0.73) | (0.69) | (0.70) |
| 4th grade residual gain | -0.26 | -0.23 | -0.24 | -0.25 | -0.27 | -0.25 |
| | (0.63) | (0.64) | (0.64) | (0.66) | (0.64) | (0.64) |
| 5th grade residual gain | -0.10 | -0.05 | -0.05 | -0.03 | 0.04 | -0.04 |
| | (0.65) | (0.68) | (0.67) | (0.69) | (0.64) | (0.67) |
| *Mathematics* | | | | | | |
| 3rd grade *z*-score | -0.68 | -0.65 | -0.63 | -0.58 | -0.56 | -0.62 |
| | (0.95) | (0.94) | (0.95) | (0.97) | (0.94) | (0.95) |
| 4th grade *z*-score | -0.79 | -0.73 | -0.71 | -0.63 | -0.64 | -0.70 |
| | (0.91) | (0.96) | (0.99) | (0.88) | (0.96) | (0.94) |
| 5th grade *z*-score | -0.70 | -0.67 | -0.56 | -0.52 | -0.49 | -0.59 |
| | (0.92) | (0.96) | (0.89) | (0.90) | (0.96) | (0.93) |
| 4th grade *z*-score gain | -0.11 | -0.08 | -0.08 | -0.05 | -0.08 | -0.08 |
| | (0.71) | (0.67) | (0.67) | (0.65) | (0.58) | (0.66) |
| 5th grade *z*-score gain | 0.09 | 0.06 | 0.15 | 0.11 | 0.15 | 0.11 |
| | (0.66) | (0.66) | (0.64) | (0.60) | (0.58) | (0.63) |
| 4th grade residual gain | -0.24 | -0.20 | -0.20 | -0.16 | -0.17 | -0.19 |
| | (0.66) | (0.64) | (0.63) | (0.58) | (0.56) | (0.62) |
| 5th grade residual gain | -0.03 | -0.05 | 0.04 | 0.02 | 0.05 | 0.00 |
| | (0.62) | (0.62) | (0.59) | (0.57) | (0.56) | (0.60) |

Table 7

*z-Score Means and Standard Deviations for Students Exiting Out of Special Education by Cohort and Across All Cohorts*

| | Cohort year | | | | | |
|---|---|---|---|---|---|---|
| | 97-99 | 98-00 | 99-01 | 00-02 | 01-03 | All |
| *Reading comprehension* | | | | | | |
| 3rd grade *z*-score | -0.37 | -0.35 | -0.35 | -0.25 | -0.33 | -0.33 |
| | (0.95) | (0.93) | (0.93) | (0.92) | (0.94) | (0.93) |
| 4th grade *z*-score | -0.37 | -0.36 | -0.36 | -0.25 | -0.30 | -0.33 |
| | (0.95) | (0.95) | (0.94) | (0.93) | (0.94) | (0.94) |
| 5th grade *z*-score | -0.37 | -0.34 | -0.38 | -0.26 | -0.28 | -0.32 |
| | (1.00) | (0.99) | (1.02) | (0.99) | (0.94) | (0.99) |
| 4th grade *z*-score gain | 0.00 | -0.01 | 0.00 | 0.01 | 0.03 | 0.01 |
| | (0.61) | (0.64) | (0.63) | (0.59) | (0.65) | (0.62) |
| 5th grade *z*-score gain | 0.00 | 0.04 | -0.01 | -0.01 | 0.03 | 0.01 |
| | (0.60) | (0.61) | (0.64) | (0.61) | (0.63) | (0.62) |
| 4th grade residual gain | -0.06 | -0.07 | -0.07 | -0.04 | -0.03 | -0.05 |
| | (0.58) | (0.61) | (0.60) | (0.56) | (0.61) | (0.59) |
| 5th grade residual gain | -0.06 | -0.02 | -0.08 | -0.06 | -0.03 | -0.05 |
| | (0.59) | (0.59) | (0.64) | (0.60) | (0.60) | (0.60) |
| *Mathematics* | | | | | | |
| 3rd grade *z*-score | -0.27 | -0.28 | -0.28 | -0.18 | -0.25 | -0.25 |
| | (0.92) | (0.94) | (0.91) | (0.92) | (0.88) | (0.91) |
| 4th grade *z*-score | -0.32 | -0.31 | -0.34 | -0.23 | -0.28 | -0.29 |
| | (0.94) | (0.96) | (0.95) | (0.93) | (0.88) | (0.93) |
| 5th grade *z*-score | -0.35 | -0.34 | -0.37 | -0.27 | -0.28 | -0.32 |
| | (0.97) | (0.96) | (0.92) | (0.93) | (0.92) | (0.94) |
| 4th grade *z*-score gain | -0.04 | -0.03 | -0.06 | -0.05 | -0.03 | -0.04 |
| | (0.59) | (0.62) | (0.61) | (0.60) | (0.56) | (0.60) |
| 5th grade *z*-score gain | -0.04 | -0.03 | -0.04 | -0.03 | 0.01 | -0.02 |
| | (0.57) | (0.57) | (0.58) | (0.50) | (0.54) | (0.55) |
| 4th grade residual gain | -0.09 | -0.08 | -0.11 | -0.09 | -0.08 | -0.09 |
| | (0.56) | (0.59) | (0.59) | (0.57) | (0.53) | (0.57) |
| 5th grade residual gain | -0.09 | -0.08 | -0.09 | -0.07 | -0.03 | -0.07 |
| | (0.55) | (0.54) | (0.54) | (0.48) | (0.53) | (0.53) |

Table 8

*z-Score Means and Standard Deviations for Students Consistently in Special Education Programs by Cohort and Across All Cohorts*

| | Cohort year | | | | | |
|---|---|---|---|---|---|---|
| | 97-99 | 98-00 | 99-01 | 00-02 | 01-03 | All |
| *Reading comprehension* | | | | | | |
| 3rd grade *z*-score | -1.00 | -1.00 | -0.88 | -0.83 | -0.84 | -0.91 |
| | (0.88) | (0.92) | (0.93) | (0.95) | (0.96) | (0.93) |
| 4th grade *z*-score | -0.99 | -0.96 | -0.83 | -0.80 | -0.80 | -0.87 |
| | (0.93) | (0.92) | (0.93) | (0.94) | (0.97) | (0.94) |
| 5th grade *z*-score | -0.97 | -0.95 | -0.83 | -0.79 | -0.75 | -0.86 |
| | (1.00) | (1.00) | (1.02) | (1.03) | (0.98) | (1.01) |
| 4th grade *z*-score gain | 0.00 | 0.03 | 0.04 | 0.01 | 0.01 | 0.02 |
| | (0.65) | (0.68) | (0.67) | (0.67) | (0.67) | (0.67) |
| 5th grade *z*-score gain | 0.02 | 0.01 | -0.02 | -0.02 | 0.08 | 0.01 |
| | (0.63) | (0.66) | (0.66) | (0.68) | (0.67) | (0.66) |
| 4th grade residual gain | -0.17 | -0.14 | -0.12 | -0.14 | -0.15 | -0.15 |
| | (0.62) | (0.64) | (0.63) | (0.63) | (0.63) | (0.63) |
| 5th grade residual gain | -0.14 | -0.15 | -0.16 | -0.16 | -0.07 | -0.14 |
| | (0.61) | (0.64) | (0.64) | (0.66) | (0.62) | (0.64) |
| *Mathematics* | | | | | | |
| 3rd grade *z*-score | -0.75 | -0.69 | -0.58 | -0.55 | -0.54 | -0.62 |
| | (0.96) | (0.93) | (0.93) | (0.92) | (0.87) | (0.92) |
| 4th grade *z*-score | -0.80 | -0.78 | -0.68 | -0.60 | -0.62 | -0.70 |
| | (0.94) | (0.94) | (0.92) | (0.83) | (0.86) | (0.90) |
| 5th grade *z*-score | -0.86 | -0.83 | -0.66 | -0.66 | -0.63 | -0.73 |
| | (0.97) | (0.96) | (0.85) | (0.85) | (0.90) | (0.91) |
| 4th grade *z*-score gain | -0.06 | -0.10 | -0.11 | -0.05 | -0.09 | -0.08 |
| | (0.67) | (0.67) | (0.66) | (0.62) | (0.56) | (0.64) |
| 5th grade *z*-score gain | -0.05 | -0.06 | 0.01 | -0.06 | -0.01 | -0.03 |
| | (0.63) | (0.61) | (0.58) | (0.53) | (0.55) | (0.58) |
| 4th grade residual gain | -0.20 | -0.22 | -0.21 | -0.16 | -0.18 | -0.19 |
| | (0.62) | (0.63) | (0.62) | (0.56) | (0.53) | (0.59) |
| 5th grade residual gain | -0.17 | -0.18 | -0.09 | -0.14 | -0.10 | -0.14 |
| | (0.60) | (0.58) | (0.54) | (0.51) | (0.53) | (0.56) |

Table 9

*z-Score Means and Standard Deviations for Students Consistently in General Education Programs by Cohort and Across All Cohorts*

| | Cohort year | | | | | |
|---|---|---|---|---|---|---|
| | 97-99 | 98-00 | 99-01 | 00-02 | 01-03 | All |
| *Reading comprehension* | | | | | | |
| 3rd grade *z*-score | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 |
| | (0.88) | (0.87) | (0.88) | (0.88) | (0.87) | (0.88) |
| 4th grade *z*-score | 0.19 | 0.18 | 0.18 | 0.18 | 0.16 | 0.18 |
| | (0.90) | (0.91) | (0.91) | (0.91) | (0.92) | (0.91) |
| 5th grade *z*-score | 0.17 | 0.17 | 0.16 | 0.16 | 0.18 | 0.17 |
| | (0.91) | (0.91) | (0.92) | (0.92) | (0.91) | (0.91) |
| 4th grade *z*-score gain | -0.03 | -0.04 | -0.04 | -0.04 | -0.05 | -0.04 |
| | (0.55) | (0.56) | (0.58) | (0.58) | (0.59) | (0.57) |
| 5th grade *z*-score gain | -0.02 | -0.02 | -0.02 | -0.02 | 0.02 | -0.01 |
| | (0.54) | (0.55) | (0.57) | (0.58) | (0.59) | (0.57) |
| 4th grade residual gain | 0.00 | 0.00 | 0.00 | 0.00 | -0.01 | 0.00 |
| | (0.53) | (0.54) | (0.56) | (0.56) | (0.57) | (0.55) |
| 5th grade residual gain | 0.01 | 0.01 | 0.01 | 0.01 | 0.05 | 0.02 |
| | (0.52) | (0.53) | (0.54) | (0.55) | (0.55) | (0.54) |
| *Mathematics* | | | | | | |
| 3rd grade *z*-score | 0.21 | 0.20 | 0.20 | 0.20 | 0.18 | 0.20 |
| | (0.87) | (0.89) | (0.89) | (0.89) | (0.93) | (0.90) |
| 4th grade *z*-score | 0.18 | 0.18 | 0.18 | 0.16 | 0.15 | 0.17 |
| | (0.92) | (0.92) | (0.92) | (0.96) | (0.96) | (0.94) |
| 5th grade *z*-score | 0.17 | 0.17 | 0.15 | 0.15 | 0.17 | 0.16 |
| | (0.93) | (0.93) | (0.97) | (0.97) | (0.94) | (0.95) |
| 4th grade *z*-score gain | -0.03 | -0.02 | -0.02 | -0.04 | -0.03 | -0.03 |
| | (0.57) | (0.57) | (0.56) | (0.58) | (0.56) | (0.57) |
| 5th grade *z*-score gain | 0.00 | -0.01 | -0.03 | -0.01 | 0.02 | -0.01 |
| | (0.53) | (0.53) | (0.54) | (0.53) | (0.53) | (0.53) |
| 4th grade residual gain | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 |
| | (0.55) | (0.54) | (0.54) | (0.57) | (0.54) | (0.55) |
| 5th grade residual gain | 0.02 | 0.02 | 0.00 | 0.01 | 0.04 | 0.02 |
| | (0.51) | (0.51) | (0.52) | (0.51) | (0.50) | (0.51) |

Table 10

*z-Score Means and Standard Deviations for Students Entering Into Special Education by Disability (Across All Cohorts)*

| | SLD | SLI | BED | OHI |
|---|---|---|---|---|
| *Reading comprehension* | | | | |
| 3rd grade *z*-score | -0.80 | -0.21 | -0.48 | -0.53 |
| | (0.91) | (0.97) | (0.88) | (0.90) |
| 4th grade *z*-score | -0.93 | -0.27 | -0.71 | -0.65 |
| | (0.93) | (1.01) | (0.93) | (0.93) |
| 5th grade *z*-score | -0.75 | -0.29 | -0.69 | -0.54 |
| | (0.98) | (1.06) | (0.93) | (0.93) |
| 4th grade *z*-score gain | -0.13 | -0.05 | -0.22 | -0.12 |
| | (0.67) | (0.61) | (0.73) | (0.71) |
| 5th grade *z*-score gain | 0.15 | -0.02 | -0.01 | 0.10 |
| | (0.71) | (0.61) | (0.79) | (0.71) |
| 4th grade residual gain | -0.28 | -0.09 | -0.31 | -0.22 |
| | (0.64) | (0.59) | (0.69) | (0.67) |
| 5th grade residual gain | -0.01 | -0.07 | -0.13 | -0.01 |
| | (0.67) | (0.59) | (0.74) | (0.67) |
| *Mathematics* | | | | |
| 3rd grade *z*-score | -0.68 | -0.22 | -0.51 | -0.62 |
| | (0.90) | (0.99) | (0.98) | (0.91) |
| 4th grade *z*-score | -0.75 | -0.22 | -0.76 | -0.72 |
| | (0.89) | (0.97) | (0.90) | (0.92) |
| 5th grade *z*-score | -0.58 | -0.21 | -0.80 | -0.63 |
| | (0.89) | (1.01) | (0.91) | (0.89) |
| 4th grade *z*-score gain | -0.07 | -0.01 | -0.25 | -0.09 |
| | (0.66) | (0.60) | (0.74) | (0.68) |
| 5th grade *z*-score gain | 0.16 | 0.01 | -0.05 | 0.08 |
| | (0.63) | (0.54) | (0.70) | (0.65) |
| 4th grade residual gain | -0.20 | -0.03 | -0.03 | -0.20 |
| | (0.61) | (0.56) | (0.67) | (0.63) |
| 5th grade residual gain | 0.05 | -0.02 | -0.16 | -0.02 |
| | (0.59) | (0.53) | (0.66) | (0.61) |

Table 11

*z-Score Means and Standard Deviations for Students Exiting Out of Special Education by Disability (Across All Cohorts)*

| | SLD | SLI | BED | OHI |
|---|---|---|---|---|
| *Reading comprehension* | | | | |
| 3rd grade *z*-score | -0.49 | -0.17 | -0.52 | -0.14 |
| | (0.94) | (0.89) | (0.91) | (0.92) |
| 4th grade *z*-score | -0.45 | -0.20 | -0.39 | -0.26 |
| | (0.97) | (0.91) | (0.89) | (0.88) |
| 5th grade *z*-score | -0.44 | -0.19 | -0.44 | -0.24 |
| | (1.00) | (0.96) | (0.98) | (0.99) |
| 4th grade *z*-score gain | 0.05 | -0.03 | 0.13 | -0.11 |
| | (0.65) | (0.58) | (0.64) | (0.68) |
| 5th grade *z*-score gain | 0.02 | 0.01 | -0.02 | 0.02 |
| | (0.64) | (0.59) | (0.63) | (0.69) |
| 4th grade residual gain | -0.04 | -0.06 | 0.04 | -0.14 |
| | (0.62) | (0.56) | (0.60) | (0.63) |
| 5th grade residual gain | -0.06 | -0.02 | -0.09 | -0.02 |
| | (0.62) | (0.57) | (0.61) | (0.67) |
| *Mathematics* | | | | |
| 3rd grade *z*-score | -0.33 | -0.14 | -0.47 | -0.19 |
| | (0.89) | (0.92) | (0.86) | (0.91) |
| 4th grade *z*-score | -0.39 | -0.17 | -0.54 | -0.21 |
| | (0.91) | (0.93) | (0.84) | (0.90) |
| 5th grade *z*-score | -0.46 | -0.16 | -0.45 | -0.37 |
| | (0.92) | (0.94) | (0.86) | (0.92) |
| 4th grade *z*-score gain | -0.06 | -0.03 | -0.07 | -0.01 |
| | (0.62) | (0.56) | (0.50) | (0.64) |
| 5th grade *z*-score gain | -0.06 | 0.01 | 0.09 | -0.17 |
| | (0.56) | (0.53) | (0.59) | (0.64) |
| 4th grade residual gain | -0.12 | -0.06 | -0.16 | -0.04 |
| | (0.59) | (0.54) | (0.56) | (0.60) |
| 5th grade residual gain | -0.12 | -0.01 | 0.01 | -0.20 |
| | (0.53) | (0.51) | (0.56) | (0.61) |

Table 12

*EOG Test Participation Rates by Cohort and Across All Cohorts*

| | Cohort year | | | | | |
|---|---|---|---|---|---|---|
| | 97-99 | 98-00 | 99-01 | 00-02 | 01-03 | All |
| *Students Entering Special Education* | | | | | | |
| 3rd grade reading | 97.72% | 97.68% | 98.15% | 97.72% | 97.69% | 97.80% |
| 4th grade reading | 98.07% | 98.20% | 98.34% | 98.45% | 97.92% | 98.20% |
| 5th grade reading | 91.51% | 92.14% | 88.94% | 88.08% | 95.76% | 91.20% |
| 3rd grade math | 97.72% | 97.74% | 98.34% | 97.79% | 97.69% | 97.87% |
| 4th grade math | 98.14% | 98.45% | 98.53% | 98.53% | 98.15% | 98.37% |
| 5th grade math | 92.80% | 93.49% | 91.69% | 92.35% | 96.99% | 93.37% |
| *Students Exiting Special Education* | | | | | | |
| 3rd grade reading | 96.30% | 94.95% | 95.62% | 96.67% | 96.96% | 96.07% |
| 4th grade reading | 96.87% | 95.73% | 96.52% | 97.63% | 96.26% | 96.60% |
| 5th grade reading | 96.11% | 96.50% | 96.33% | 97.72% | 96.78% | 96.69% |
| 3rd grade math | 96.67% | 96.60% | 96.52% | 97.72% | 97.56% | 97.03% |
| 4th grade math | 96.96% | 96.80% | 97.00% | 97.63% | 96.70% | 97.01% |
| 5th grade math | 96.20% | 96.70% | 96.52% | 98.01% | 96.70% | 96.82% |
| *Students Always in Special Education* | | | | | | |
| 3rd grade reading | 76.82% | 71.73% | 70.82% | 71.27% | 71.86% | 72.41% |
| 4th grade reading | 72.49% | 69.58% | 68.12% | 67.39% | 66.75% | 68.46% |
| 5th grade reading | 69.20% | 66.92% | 61.70% | 61.42% | 68.63% | 65.50% |
| 3rd grade math | 78.55% | 75.06% | 74.47% | 75.10% | 75.73% | 75.73% |
| 4th grade math | 75.52% | 73.25% | 72.30% | 72.31% | 75.52% | 72.92% |
| 5th grade math | 72.01% | 69.92% | 65.40% | 66.54% | 72.01% | 68.88% |
| *Students Always in General Education* | | | | | | |
| 3rd grade reading | 99.24% | 99.26% | 99.20% | 99.05% | 99.03% | 99.16% |
| 4th grade reading | 99.59% | 99.60% | 99.55% | 99.54% | 99.48% | 99.55% |
| 5th grade reading | 99.51% | 99.53% | 99.49% | 99.48% | 99.44% | 99.49% |
| 3rd grade math | 99.30% | 99.30% | 99.24% | 99.07% | 99.05% | 99.19% |
| 4th grade math | 99.60% | 99.61% | 99.55% | 99.56% | 99.49% | 99.56% |
| 5th grade math | 99.52% | 99.52% | 99.50% | 99.48% | 99.45% | 99.49% |

*Creating Differenced Scores*

*Gain scores.* Differenced scores were calculated to compare growth in reading and mathematics achievement scores over time. In order to test the first hypothesis, differenced scores were created using fourth grade and fifth grade gain scores. Specifically, differenced scores were calculated by subtracting fourth grade (Year 1) $z$-score gain from fifth grade (Year 2) $z$-score gain. This change in growth was then used as the dependent variable in the regression analyses that would test the first hypothesis.

*Residual gain scores.* Differenced scores were also calculated using residual gain scores in order to test the second and third hypotheses. The creation of these differenced residual gain scores also involved subtracting fourth grade (Year 1) residual gain from fifth grade (Year 2) residual gain. This change in growth was then used as the dependent variable in the regression analyses that tested the second and third hypotheses.

*Accommodation changes.* A similar process was also used to create differenced scores for students' use of accommodations. First, dichotomous variables were created to indicate whether students did or did not use accommodations during the third, fourth, and fifth grades. Second, the change in accommodation use was calculated for each year (e.g., fourth grade change equaled students' fourth grade accommodation use minus their third grade accommodation use). Finally, the fourth grade (Year 1) change in accommodations use was subtracted from the fifth grade (Year 2) change in accommodations use. This differenced accommodations variable was used as an independent variable in the regression analyses that tested the third hypothesis. Table 13 presents the means for students' differenced gain scores, differenced residual gain scores, and differenced accommodations information by group.

Table 13

*Mean Differenced Gain Scores, Differenced Residual Gain Scores, and Differenced Accommodations Use by Group*

| | Student Group | | | |
|---|---|---|---|---|
| | Entered Special Education | Exited Special Education | Always in Special Education | Always in General Education |
| Mean differenced gain in reading | 0.16 | 0.00 | -0.02 | 0.02 |
| Mean differenced gain in math | 0.13 | 0.02 | 0.03 | 0.02 |
| Mean differenced residual gain in reading | 0.20 | 0.00 | 0.00 | 0.02 |
| Mean differenced residual gain in math | 0.20 | 0.02 | 0.06 | 0.01 |
| Mean differenced accommodations use | 0.56 | -0.31 | -0.04 | 0.00 |

*Note.* All values for mean differenced gain and mean differenced residual gain scores are reported in *z*-score units based on students' grade and cohort.

*Special education status changes.* This same process was also used to create differenced scores for students' special education status. First, dichotomous variables were created to indicate students' special education status (i.e., general or special education) during the third, fourth, and fifth grades. Second, the change in special education status was calculated for each year (e.g., fourth grade change equaled students' fourth grade special education status minus their third grade special education status). Finally, the fourth grade (Year 1) change in special education status was subtracted from the fifth grade (Year 2) change in special education status. This differenced special education variable was used as an independent variable in the regression analysis that would test the first hypothesis. Additionally, students who entered and exited special education programs during the fifth grade were examined separately to estimate the differing effects for each group of students. These separate effects were the only effects estimated in the analyses used to test the second and third hypotheses, as Hanushek et al. (1998) had noted that the best estimate of special education's effect were coefficients for students entering special education, because achievement scores for exiters would likely still reflect some residual effects of receiving special education services in prior years.

*Adding School-Level Demographic Data*

School-level data are recorded by year in the NCERDC database in separate files from the student-level data. The school-level demographic variables used by Hanushek et al. (1998) in their analyses (the percent of students with free or reduced price lunch, the percent of Black students, and the percent of Hispanic students) were taken from the school-level files and merged with the student-level data, such that the appropriate school-level demographic variables were available for each student for his or her fifth grade year.

For 3,524 students, relevant school-level data were not available within the school-level files; therefore, these students were excluded from analyses. School lunch data were missing for an additional 23,021 students. For 22,145 of these students, school lunch data were able to be interpolated from the student-level data set. Missing school lunch data were estimated at the student level from the overall population of 515,095 students. When a minimum of 30 students per year per school was available, an estimated value for the percent of students receiving a free or reduced price lunch was calculated at the student level. The remaining 876 students with missing school lunch data were excluded from analyses.

### *Regression Analyses to Test the Hypotheses*

#### *Hypothesis 1*

The first hypothesis stated that receipt of special education services would be a positive predictor of academic gain in the areas of reading comprehension and mathematics. Fixed effects ordinary least squares regression analyses were computed to test Hypothesis 1. These analyses assessed the effectiveness of special education programs by determining whether students with disabilities (particularly those who entered special education programs) demonstrated greater gains on achievement tests when enrolled in special education programs than when enrolled in general education programs, relative to students who did not transition in and out of special education programs.

The regression formula was similar to the one used by Hanushek et al. (1998). The analysis was performed a total of 15 times with both EOG-Reading and EOG-Math scores. The initial three analyses included all categories of special education students and were the primary analyses to test Hypothesis 1. The remaining analyses were supplemental, and examined outcomes for four different categories of exceptionality. Specifically, the first regression looked at gains for special education entrants and exiters together. The second regression looked at gains only for students who entered special education programs. The third regression looked at gains only for students who exited special education programs. After these three initial analyses to test Hypothesis 1 were completed for all special education students, they were rerun (a) for students with learning disabilities, (b) for students with speech-language impairments, (c) for students with behavioral-emotional disabilities, and (d) for students with other health impairments. In all cases, students who remained in general education or remained in special education were included in the regression analyses as the comparison group. Each of these 15 regressions was conducted as Hanushek et al. conducted it, after excluding the bottom 1% of fourth and fifth grade test scores and the top and bottom 1% of first-level differenced scores indicating test score gain (their procedure for adjusting for regression toward the mean).

It was expected that students' special education status (e.g., entered or exited special education during fifth grade) would be a significant, positive predictor of their change in gain scores from the fourth to the fifth grade. In coding special education status for the analyses in which entrants and exiters were considered together, a negative value (-1) for special education status was used for students who exited from special education in the fifth grade and a positive value for special education status (+1) was used for students who entered into special education in the fifth grade. Thus, the two groups could be combined and special education status would have a positive *B* coefficient if students' gains were larger while in special education regardless of when they received special education services. However, in coding special education status for the analyses in which entrants and exiters were considered separately, special education entry and special education exit were labeled as two separate variables with values of +1 to indicate a transition and 0 to indicate that a student did not change programs. Because differenced gain scores (and differenced residual gain scores) were always created by subtracting fourth grade gain from fifth grade gain, special education entry would have a positive *B* coefficient if students' gain scores were larger while in special education programs and special education exit would have a negative *B* coefficient if students' gain scores were larger while in special education programs.

Overall, the results provided support for the first hypothesis. Results of the first set of regressions (in which all students with any type of disability were included) revealed that students' special education status was a significant predictor of increased growth in EOG Test scores in both reading and math after controlling for the effects of school-level characteristics and cohort. When entrants and exiters were examined separately, the effects of special education were significant for those who entered special education programs, but not for those who exited special education programs. That is, students who entered special education programs had higher gain-in-gain scores than peers who did not change their special education status between grades (see Tables 14 and 15). Sample sizes for regressions involving reading and math scores varied slightly due to missing data when student scores were available for one test and not the other.

Table 14

*Regression Analyses Predicting Differenced Gain Scores for All Transitioning Students*

| Predictors | Reading *B* | Math *B* |
|---|---|---|
| % FRL | 0.06*** | -0.05*** |
| % Black | 0.01 | 0.09*** |
| % Hispanic | 0.04 | 0.10*** |
| Cohort1 | -0.04*** | -0.04*** |
| Cohort2 | -0.04*** | -0.06*** |
| Cohort3 | -0.04*** | -0.08*** |
| Cohort4 | -0.04*** | -0.03*** |
| Special education status | 0.08*** | 0.06*** |
| | | |
| $R^2$ | 0.0010 | 0.0018 |
| *Sample size (N)* | 379,352 | 382,112 |

*Note.* Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 15

*Regression Analyses Predicting Differenced Gain Scores for Students Entering or Exiting Special Education Programs*

| Predictors | Reading *B* | Math *B* |
|---|---|---|
| *Special education entry* | | |
| % FRL | 0.06*** | -0.05*** |
| % Black | 0.01 | 0.09*** |
| % Hispanic | 0.04 | 0.10*** |
| Cohort1 | -0.04*** | -0.04*** |
| Cohort2 | -0.04*** | -0.06*** |
| Cohort3 | -0.04*** | -0.08*** |
| Cohort4 | -0.04*** | -0.03*** |
| Special education entry | 0.14*** | 0.11*** |
| $R^2$ | 0.0011 | 0.0019 |
| *Sample size (N)* | 374,684 | 377,373 |
| | | |
| *Special education exit* | | |
| % FRL | 0.06*** | -0.05*** |
| % Black | 0.01 | 0.09*** |
| % Hispanic | 0.05* | 0.11*** |
| Cohort1 | -0.04*** | -0.04*** |
| Cohort2 | -0.04** | -0.06*** |
| Cohort3 | -0.04*** | -0.08*** |
| Cohort4 | -0.04*** | -0.03*** |
| Special education exit | -0.02 | 0.00 |
| $R^2$ | 0.0008 | 0.0016 |
| *Sample size (N)* | 373,694 | 376,280 |

*Note*. Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note*. A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

These analyses were rerun separately for students with SLD, SLI, BED, and OHI in the areas of reading and mathematics (see Tables 16 and 17). For students with SLD or OHI, special education status continued to be a significant predictor of reading and mathematics differenced gain scores when special education status was examined for all transitioning students, or for those who entered special education programs. For students with BED, the effectiveness of special education was demonstrated in the area of reading, but not mathematics. For students with SLI, special education status was not a significant predictor of changes in achievement scores in either academic subject.

Table 16

*Regression Analysis Predicting Differenced Gain Scores for Students With SLD, SLI, BED, or OHI in the Area of Reading Comprehension*

| Predictors | SLD *B* | SLI *B* | BED *B* | OHI *B* |
|---|---|---|---|---|
| *Special education status* | | | | |
| % FRL | 0.06*** | 0.06*** | 0.06*** | 0.07*** |
| % Black | 0.01 | 0.02* | 0.02 | 0.02* |
| % Hispanic | 0.06* | 0.07** | 0.07** | 0.07** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort4 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Special education status | 0.14*** | -0.02 | 0.15*** | 0.10*** |
| $R^2$ | 0.0011 | 0.0008 | 0.0008 | 0.0008 |
| *Sample size (N)* | 365,585 | 349,861 | 347,516 | 350,021 |
| | | | | |
| *Special education entry* | | | | |
| % FRL | 0.06*** | 0.06*** | 0.06*** | 0.07*** |
| % Black | 0.01 | 0.02* | 0.02 | 0.02* |
| % Hispanic | 0.06* | 0.07** | 0.07** | 0.07** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort4 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Special education entry | 0.20*** | -0.02 | 0.15** | 0.13*** |
| $R^2$ | 0.0012 | 0.0008 | 0.0008 | 0.0009 |
| *Sample size (N)* | 363,599 | 347,661 | 347,394 | 349,845 |
| | | | | |
| *Special education exit* | | | | |
| % FRL | 0.07*** | 0.06*** | 0.06*** | 0.07*** |
| % Black | 0.01 | 0.02* | 0.02* | 0.02 |
| % Hispanic | 0.06* | 0.07** | 0.07** | 0.07** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort4 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Special education exit | -0.05** | 0.02 | -0.14 | 0.07 |
| $R^2$ | 0.0008 | 0.0008 | 0.0008 | 0.0008 |
| *Sample size (N)* | 362,391 | 349,165 | 347,137 | 349,070 |

*Note.* Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note.* A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 17

*Regression Analysis Predicting Differenced Gain Scores for Students With SLD, SLI, BED, or OHI in the Area of Mathematics*

| Predictors | SLD *B* | SLI *B* | BED *B* | OHI *B* |
|---|---|---|---|---|
| *Special education status* | | | | |
| % FRL | -0.05*** | -0.05*** | -0.05*** | -0.05*** |
| % Black | 0.10*** | 0.10*** | 0.10*** | 0.10*** |
| % Hispanic | 0.11*** | 0.12*** | 0.12*** | 0.12*** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.06*** | -0.06*** | -0.06*** | -0.06*** |
| Cohort3 | -0.08*** | -0.08*** | -0.08*** | -0.08*** |
| Cohort4 | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| Special education status | 0.09*** | -0.02 | 0.03 | 0.10*** |
| $R^2$ | 0.0019 | 0.0018 | 0.0018 | 0.0018 |
| *Sample size (N)* | 368,303 | 349,849 | 347,551 | 350,003 |
| *Special education entry* | | | | |
| % FRL | -0.05*** | -0.05*** | -0.05*** | -0.05*** |
| % Black | 0.10*** | 0.10*** | 0.10*** | 0.10*** |
| % Hispanic | 0.11*** | 0.12*** | 0.12*** | 0.12*** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.06*** | -0.06*** | -0.06*** | -0.06*** |
| Cohort3 | -0.08*** | -0.08*** | -0.08*** | -0.08*** |
| Cohort4 | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| Special education entry | 0.15*** | -0.03 | 0.08 | 0.09** |
| $R^2$ | 0.0020 | 0.0020 | 0.0018 | 0.0018 |
| *Sample size (N)* | 366,238 | 347,651 | 347,426 | 349,832 |
| *Special education exit* | | | | |
| % FRL | -0.05*** | -0.05*** | -0.05*** | -0.05*** |
| % Black | 0.10*** | 0.10*** | 0.10*** | 0.10*** |
| % Hispanic | 0.11*** | 0.12*** | 0.12*** | 0.12*** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.06*** | -0.06*** | -0.06*** | -0.06*** |
| Cohort3 | -0.08*** | -0.08*** | -0.08*** | -0.08*** |
| Cohort4 | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| Special education exit | 0.00 | 0.01 | 0.13 | -0.16* |
| $R^2$ | 0.0017 | 0.0018 | 0.0018 | 0.0018 |
| *Sample size (N)* | 364,886 | 349,154 | 347,149 | 349,071 |

*Note.* Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note.* A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

In sum, the results of the present study replicated the findings of Hanushek et al. (1998) and provided support for the first hypothesis. When all students who transitioned in and out of special education programs were included in the regression analysis, special education status was a significant predictor of differenced *z*-score gain on the EOG-Reading and EOG-Math Tests. This finding remained true when considering students with SLD or OHI. Significant results were found for students with BED, but only in the area of reading. When students who entered and exited special education programs were considered separately, students who entered special education programs demonstrated greater achievement gains when involved in special education programs both when students with all types of disabilities were considered together and when students with SLD or OHI were considered separately. Again, students with BED only demonstrated significant effects in the area of reading. Students who exited special education programs demonstrated a less consistent pattern of results: students with SLD demonstrated effects in the area of reading and students with OHI demonstrated effects in the area of math. Interestingly, the school-level and cohort control variables were significant predictors of achievement differences in nearly every analysis.

*Hypothesis 2*

The second hypothesis stated that the advantages of special education over general education placement would continue to exist after controlling for regression to the mean. Fixed effects ordinary least squares regression analyses similar to the second and third analyses used in Hypothesis 1 were conducted to test Hypothesis 2. The initial regression in Hypothesis 1 (where special education entrants and exiters were considered together) was omitted in this set of analyses, as well as those for Hypothesis 3. This strategy was in keeping with Hanushek et al.'s (1998) concern that gains for special education exiters might reflect residual effects of special education and their initial finding, confirmed in the analyses for Hypothesis 1, that effects for special education entrants and exiters differed.

As with Hypothesis 1, the new regression analyses were first run for all students with disabilities, and then for each of the four largest disability groups (SLD, SLI, BED, and OHI), with both EOG-Reading and EOG-Math scores. However, the dependent variable for each equation in this set of analyses was students' differenced residual gain scores for reading and math. Even after controlling for regression to the mean, it was expected that students who entered special education in the fifth grade would demonstrate significantly greater residual gain when they were enrolled in special education programs than when they were enrolled in general education programs. It was expected that this effect would be less likely to be significant for students who exited special education programs in the fifth grade.

As predicted, students' entry into special education remained a significant predictor of differenced residual gain scores in the areas of both reading and math when all students with disabilities were considered as a group, as well as when students with SLD, BED, and OHI were considered separately. Students' exit from special education was a significant predictor of differenced residual gain scores in the area of reading for students with SLD, in the area of math for students with OHI, and in the area of math for students with BED (though in the opposite direction). Special education status was not a significant predictor of scores for students with SLI. Parameter estimates (*B*) for each predictor are provided in Tables 18, 19, and 20.

Table 18

*Regression Analyses Predicting Differenced Residual Gain Scores for Students Entering or Exiting Special Education Programs*

| Predictors | Reading *B* | Math *B* |
|---|---|---|
| *Special education entry* | | |
| % FRL | 0.05*** | -0.05*** |
| % Black | 0.02** | 0.12*** |
| % Hispanic | 0.08*** | 0.16*** |
| Cohort1 | -0.04*** | -0.02*** |
| Cohort2 | -0.05*** | -0.04*** |
| Cohort3 | -0.06*** | -0.05*** |
| Cohort4 | -0.05*** | -0.03*** |
| Special education entry | 0.19*** | 0.18*** |
| $R^2$ | 0.0016 | 0.0020 |
| *Sample size (N)* | 388,669 | 391,221 |
| | | |
| *Special education exit* | | |
| % FRL | 0.05*** | -0.05*** |
| % Black | 0.03*** | 0.12*** |
| % Hispanic | 0.09*** | 0.16*** |
| Cohort1 | -0.04*** | -0.02*** |
| Cohort2 | -0.05** | -0.04*** |
| Cohort3 | -0.05*** | -0.05*** |
| Cohort4 | -0.05*** | -0.03*** |
| Special education exit | -0.01 | 0.00 |
| $R^2$ | 0.0010 | 0.0013 |
| *Sample size (N)* | 387,260 | 389,713 |

*Note*. Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note*. A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 19

*Regression Analysis Predicting Differenced Residual Gain Scores for Students With SLD, SLI, BED, or OHI in the Area of Reading*

| | SLD | SLI | BED | OHI |
|---|---|---|---|---|
| Predictors | *B* | *B* | *B* | *B* |
| *Special education entry* | | | | |
| % FRL | 0.05*** | 0.05*** | 0.05*** | 0.05*** |
| % Black | 0.03*** | 0.03*** | 0.03*** | 0.03*** |
| % Hispanic | 0.10*** | 0.11*** | 0.12*** | 0.11*** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.05*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.05*** | -0.05*** | -0.05*** | -0.05*** |
| Cohort4 | -0.05*** | -0.04*** | -0.04*** | -0.05*** |
| Special education entry | 0.25*** | 0.00 | 0.15** | 0.19*** |
| $R^2$ | 0.0017 | 0.0010 | 0.0010 | 0.0012 |
| *Sample size (N)* | 375,983 | 357,478 | 357,452 | 360,008 |
| | | | | |
| *Special education exit* | | | | |
| % FRL | 0.05*** | 0.05*** | 0.05*** | 0.05*** |
| % Black | 0.03*** | 0.03*** | 0.03*** | 0.03*** |
| % Hispanic | 0.10*** | 0.12*** | 0.12*** | 0.11*** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.05*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.05*** | -0.05*** | -0.05*** | -0.05*** |
| Cohort4 | -0.05*** | -0.04*** | -0.04*** | -0.04*** |
| Special education exit | -0.04* | 0.02 | -0.15 | 0.11 |
| $R^2$ | 0.0010 | 0.0010 | 0.0010 | 0.0010 |
| *Sample size (N)* | 374,469 | 359,044 | 357,157 | 359,146 |

*Note*. Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note*. A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 20

*Regression Analysis Predicting Differenced Residual Gain Scores for Students With SLD, SLI, BED, or OHI in the area of Mathematics*

| | SLD | SLI | BED | OHI |
|---|---|---|---|---|
| Predictors | *B* | *B* | *B* | *B* |
| *Special education entry* | | | | |
| % FRL | -0.06*** | -0.06*** | -0.06*** | -0.06*** |
| % Black | 0.12*** | 0.13*** | 0.13*** | 0.13*** |
| % Hispanic | 0.17*** | 0.18*** | 0.18*** | 0.17*** |
| Cohort1 | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| Cohort2 | -0.04*** | -0.04*** | -0.04*** | -0.03*** |
| Cohort3 | -0.05*** | -0.06*** | -0.06*** | -0.06*** |
| Cohort4 | -0.03*** | -0.02*** | -0.02*** | -0.02*** |
| Special education entry | 0.23*** | -0.01 | 0.16*** | 0.16*** |
| $R^2$ | 0.0021 | 0.0016 | 0.0016 | 0.0016 |
| *Sample size (N)* | 378,123 | 357,637 | 357,713 | 360,260 |
| | | | | |
| *Special education exit* | | | | |
| % FRL | -0.05*** | -0.06*** | -0.06*** | -0.06*** |
| % Black | 0.12*** | 0.13*** | 0.13*** | 0.13*** |
| % Hispanic | 0.17*** | 0.18*** | 0.18*** | 0.17*** |
| Cohort1 | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| Cohort2 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.05*** | -0.06*** | -0.06*** | -0.06*** |
| Cohort4 | -0.03*** | -0.02*** | -0.02*** | -0.02*** |
| Special education exit | -0.01 | 0.03 | 0.15* | -0.17** |
| $R^2$ | 0.0014 | 0.0016 | 0.0015 | 0.0015 |
| *Sample size (N)* | 376,538 | 359,206 | 357,410 | 359,395 |

*Note.* Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note.* A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

In sum, results extend the findings of Hanushek et al. (1998) and provided support for the second hypothesis. Students' entry into special education programs was a significant predictor of their achievement score gain after the effects of regression to the mean were taken into consideration. These results were found in the areas of both reading and math when students with all disabilities were considered together, as well as when students with SLD, BED, or OHI were considered separately. Students' exit from special education programs was also a significant predictor of reading for students with SLD in the area of reading and for students with OHI in the area of math. Contrary to predictions, for students with BED, exit from special education was a predictor of achievement gain in the opposite direction that would have been expected. Once again, the school-level and cohort control variables were significant predictors of residual achievement differences in nearly every analysis.

*Hypothesis 3*

The final hypothesis stated that the advantages of special education over general education placement would continue to exist after controlling for the use of accommodations by students in special education programs. This hypothesis was tested using two different methods: (a) entering students' use of accommodations as a separate independent variable into the prediction equation and (b) running the regression analyses solely for students who did not receive accommodations on the EOG Tests.

*Controlling for accommodation use.* Fixed effects ordinary least squares regression analyses were computed to test Hypothesis 3. Similar to the previous analyses, differenced residual gain scores were used as outcome variables to continue to account for potential regression toward the mean. Additionally, this analysis accounted for the use of accommodations by students in special education programs. Again, the regression analyses were run for all students with disabilities, as well as for each of the four largest disability groups, with both EOG-Reading and EOG-Math Test scores. Even after controlling for regression to the mean, school-level variables, cohort variables, and the use of accommodations, it was expected that special education entry (and to a lesser extent, special education exit) would be a significant predictor of students' differenced residual gain scores. However, it was expected that the effect of special education programs on academic achievement outcomes would be reduced with the differenced accommodations variable entered into the regression equation. Results revealed that entry into special education programs was a significant predictor of differenced residual gain in both math and reading when all students with disabilities were included in the analyses (see Table 21). Students' use of accommodations was also a significant predictor of differenced residual gain for students entering and exiting special education programs. Students' exit from special education programs was a significant predictor of their differenced residual gain scores in the area of mathematics; however, this finding was in the opposite direction as predicted in the hypothesis.

Table 21

*Regression Analyses Predicting Differenced Residual Gain Scores After Controlling for Accommodation Use*

| Predictors | Reading *B* | Math *B* |
|---|---|---|
| *Special education entry* | | |
| % FRL | 0.05*** | -0.06*** |
| % Black | 0.02** | 0.12*** |
| % Hispanic | 0.08*** | 0.16*** |
| Cohort1 | -0.05*** | -0.02*** |
| Cohort2 | -0.05*** | -0.04*** |
| Cohort3 | -0.06*** | -0.05*** |
| Cohort4 | -0.05*** | -0.03*** |
| Accommodation use | 0.08*** | 0.10*** |
| Special education entry | 0.14*** | 0.12*** |
| $R^2$ | 0.0023 | 0.0030 |
| *Sample size (N)* | 388,669 | 391,221 |
| | | |
| *Special education exit* | | |
| % FRL | 0.05*** | -0.05*** |
| % Black | 0.03*** | 0.12*** |
| % Hispanic | 0.09*** | 0.16*** |
| Cohort1 | -0.04*** | -0.02*** |
| Cohort2 | -0.05*** | -0.04*** |
| Cohort3 | -0.05*** | -0.05*** |
| Cohort4 | -0.05*** | -0.03*** |
| Accommodation use | 0.08*** | 0.09*** |
| Special education exit | 0.01 | 0.03* |
| $R^2$ | 0.0015 | 0.0021 |
| *Sample size (N)* | 387,260 | 389,713 |

*Note.* Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note.* A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Entry into special education programs continued to be a significant predictor of differenced residual gain in both math and reading for students with LD, BED, or OHI; exit from special education programs was a significant predictor of differenced residual gain in math for students with OHI. Exit from special education programs was also a significant predictor of differenced residual gain in reading for students with OHI and for gain in math for students with SLI and BED; however, these significant findings for exiters were in the opposite direction than was hypothesized. Students' use of accommodations also tended to be a significant predictor of differenced achievement gains. The school-level and cohort control variables continued to be significant predictors of residual achievement differences in nearly every analysis. Results of these analyses are reported in Tables 22 and 23.

Table 22

*Regression Analysis Predicting Differenced Residual Gain Scores for Students With SLD, SLI, BED, or OHI in the Area of Reading After Controlling for Accommodation Use*

| | SLD | SLI | BED | OHI |
|---|---|---|---|---|
| Predictors | *B* | *B* | *B* | *B* |
| *Special education entry* | | | | |
| % FRL | 0.05*** | 0.05*** | 0.05*** | 0.05*** |
| % Black | 0.03*** | 0.03*** | 0.03*** | 0.03*** |
| % Hispanic | 0.10*** | 0.11*** | 0.12*** | 0.11*** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.05*** | -0.04*** | -0.04*** | -0.05*** |
| Cohort3 | -0.05*** | -0.05*** | -0.05*** | -0.05*** |
| Cohort4 | -0.05*** | -0.04*** | -0.04*** | -0.05*** |
| Accommodation use | 0.09*** | 0.09*** | 0.09*** | 0.09*** |
| Special education entry | 0.19*** | -0.02 | 0.10* | 0.14*** |
| $R^2$ | 0.0023 | 0.0016 | 0.0016 | 0.0017 |
| *Sample size (N)* | 375,983 | 357,478 | 357,452 | 360,008 |
| *Special education exit* | | | | |
| % FRL | 0.05*** | 0.05*** | 0.05*** | 0.05*** |
| % Black | 0.03*** | 0.03*** | 0.03*** | 0.03*** |
| % Hispanic | 0.10*** | 0.12*** | 0.12*** | 0.11*** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.05*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.05*** | -0.05*** | -0.05*** | -0.05*** |
| Cohort4 | -0.05*** | -0.04*** | -0.04*** | -0.05*** |
| Accommodation use | 0.08*** | 0.09*** | 0.09*** | 0.08*** |
| Special education exit | 0.00 | 0.03 | -0.11 | 0.15* |
| $R^2$ | 0.0016 | 0.0015 | 0.0016 | 0.0015 |
| *Sample size (N)* | 374,469 | 359,044 | 357,157 | 359,146 |

*Note.* Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note.* A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 23

*Regression Analysis Predicting Differenced Residual Gain Scores for Students With SLD, SLI, BED, or OHI in the Area of Mathematics After Controlling for Accommodation Use*

| | SLD | SLI | BED | OHI |
|---|---|---|---|---|
| Predictors | *B* | *B* | *B* | *B* |
| *Special education entry* | | | | |
| % FRL | -0.06*** | -0.06*** | -0.06*** | -0.06*** |
| % Black | 0.12*** | 0.13*** | 0.13*** | 0.13*** |
| % Hispanic | 0.17*** | 0.18*** | 0.18*** | 0.17*** |
| Cohort1 | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| Cohort2 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.05*** | -0.06*** | -0.06*** | -0.06*** |
| Cohort4 | -0.03*** | -0.02*** | -0.02*** | -0.02*** |
| Accommodation use | 0.10*** | 0.10*** | 0.10*** | 0.10*** |
| Special education entry | 0.17*** | -0.02 | 0.10* | 0.11*** |
| $R^2$ | 0.0029 | 0.0022 | 0.0022 | 0.0023 |
| *Sample size (N)* | 378,123 | 357,637 | 357,713 | 360,260 |
| | | | | |
| *Special education exit* | | | | |
| % FRL | -0.06*** | -0.06*** | -0.06*** | -0.06*** |
| % Black | 0.12*** | 0.13*** | 0.13*** | 0.13*** |
| % Hispanic | 0.17*** | 0.18*** | 0.18*** | 0.17*** |
| Cohort1 | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| Cohort2 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.05*** | -0.06*** | -0.06*** | -0.06*** |
| Cohort4 | -0.03*** | -0.02*** | -0.02*** | -0.02*** |
| Accommodation use | 0.09*** | 0.10*** | 0.10*** | 0.10*** |
| Special education exit | 0.04 | 0.04* | 0.20** | -0.12* |
| $R^2$ | 0.0021 | 0.0022 | 0.0022 | 0.0022 |
| *Sample size (N)* | 376,538 | 359,206 | 357,410 | 359,395 |

*Note.* Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note.* A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

*Excluding students who received accommodations.* In order to examine the relationship between special education status and differenced residual gain scores on EOG-Reading and Math Tests after controlling for the effects of accommodation use further, one final set of regressions were conducted.  Here, the regressions were conducted solely for students who did not receive any accommodations during the EOG Tests in the third, fourth, or fifth grades.  Results (presented in Table 24) revealed that, as hypothesized, students' entry into special education programs was a positive, significant predictor of their differenced residual scores in the area of math; for the first and only time throughout the study, entry into special education was not a significant predictor of differenced residual gain scores in the area of reading.  Although students' exit from special education programs was also a significant predictor in the area of mathematics, it was in the opposite direction that was hypothesized.  Once again, the school-level and cohort control variables were significant predictors of residual achievement differences in every analysis.  Of the students that did not receive accommodations during the third, fourth, or fifth grade EOG Tests, 98.27% were enrolled in general education programs consistently throughout those three grades.  Additionally, 0.45% entered special education programs during the fifth grade, 0.69% exited special education programs during the fifth grade, and 0.59% remained in special education programs throughout the study.

Table 24

*Regression Analyses Predicting Differenced Residual Gain Scores for Students Who Did Not Receive Accommodations*

| Predictors | Reading *B* | Math *B* |
|---|---|---|
| *Special education entry* | | |
| % FRL | 0.05*** | -0.06*** |
| % Black | 0.03*** | 0.13*** |
| % Hispanic | 0.12*** | 0.17*** |
| Cohort1 | -0.04*** | -0.02*** |
| Cohort2 | -0.04*** | -0.03*** |
| Cohort3 | -0.05*** | -0.06*** |
| Cohort4 | -0.05*** | -0.02*** |
| Special education entry | 0.04 | 0.05* |
| $R^2$ | 0.0011 | 0.0016 |
| *Sample size (N)* | 344,692 | 344,764 |
| | | |
| *Special education exit* | | |
| % FRL | 0.05*** | -0.06*** |
| % Black | 0.03*** | 0.13*** |
| % Hispanic | 0.13*** | 0.17*** |
| Cohort1 | -0.04*** | -0.02*** |
| Cohort2 | -0.04*** | -0.03*** |
| Cohort3 | -0.05*** | -0.06*** |
| Cohort4 | -0.05*** | -0.02*** |
| Special education exit | 0.02 | 0.03* |
| $R^2$ | 0.0011 | 0.0016 |
| *Sample size (N)* | 345,534 | 345,604 |

*Note*. Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note*. A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Entry into special education programs was a significant predictor of differenced residual gain scores in both math and reading for students with SLD. Exit from special education programs was also a significant predictor of differenced residual gain in math for students with SLD; however, these significant findings for exiters were in the opposite direction than was hypothesized. The school-level and cohort control variables continued to be significant predictors of residual achievement differences in nearly every analysis. Results of these analyses are reported in Tables 25 and 26.

Table 25

*Regression Analysis Predicting Differenced Residual Gain Scores for Students With SLD, SLI, BED, or OHI Without Accommodations in the Area of Reading*

| | SLD | SLI | BED | OHI |
|---|---|---|---|---|
| Predictors | *B* | *B* | *B* | *B* |
| *Special education entry* | | | | |
| % FRL | 0.05*** | 0.05*** | 0.05*** | 0.05*** |
| % Black | 0.03*** | 0.03*** | 0.03*** | 0.03*** |
| % Hispanic | 0.13*** | 0.12*** | 0.13*** | 0.13*** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.05*** | -0.05*** | -0.05*** | -0.05*** |
| Cohort4 | -0.05*** | -0.05*** | -0.05*** | -0.05*** |
| Special education entry | 0.09** | -0.03 | 0.07 | -0.03 |
| $R^2$ | 0.0011 | 0.0011 | 0.0011 | 0.0011 |
| *Sample size (N)* | 342,449 | 342,734 | 341,243 | 341,312 |
| | | | | |
| *Special education exit* | | | | |
| % FRL | 0.05*** | 0.05*** | 0.05*** | 0.05*** |
| % Black | 0.03*** | 0.03*** | 0.03*** | 0.03*** |
| % Hispanic | 0.13*** | 0.13*** | 0.13*** | 0.13*** |
| Cohort1 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort2 | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| Cohort3 | -0.05*** | -0.05*** | -0.05*** | -0.05*** |
| Cohort4 | -0.05*** | -0.05*** | -0.05 | -0.05*** |
| Special education exit | 0.00 | 0.03 | 0.25 | 0.03 |
| $R^2$ | 0.0011 | 0.0011 | 0.0011 | 0.0011 |
| *Sample size (N)* | 342,140 | 344,135 | 341,186 | 341,213 |

*Note.* Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note.* A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 26

*Regression Analysis Predicting Differenced Residual Gain Scores for Students With SLD, SLI, BED, or OHI Without Accommodations in the Area of Mathematics*

| | SLD | SLI | BED | OHI |
|---|---|---|---|---|
| Predictors | *B* | *B* | *B* | *B* |
| *Special education entry* | | | | |
| % FRL | -0.06*** | -0.06*** | -0.06*** | -0.06*** |
| % Black | 0.13*** | 0.13*** | 0.13*** | 0.13*** |
| % Hispanic | 0.17*** | 0.17*** | 0.17*** | 0.17*** |
| Cohort1 | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| Cohort2 | -0.03*** | -0.04*** | -0.04*** | -0.03*** |
| Cohort3 | -0.06*** | -0.06*** | -0.06*** | -0.06*** |
| Cohort4 | -0.03*** | -0.02*** | -0.02*** | -0.02*** |
| Special education entry | 0.08* | -0.01 | 0.18 | -0.02 |
| $R^2$ | 0.0016 | 0.0016 | 0.0016 | 0.0016 |
| *Sample size (N)* | 342,519 | 342,811 | 341,321 | 341,387 |
| | | | | |
| *Special education exit* | | | | |
| % FRL | -0.06*** | -0.06*** | -0.06*** | -0.06*** |
| % Black | 0.13*** | 0.13*** | 0.13*** | 0.13*** |
| % Hispanic | 0.17*** | 0.17*** | 0.17*** | 0.17*** |
| Cohort1 | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| Cohort2 | -0.03*** | -0.04*** | -0.03*** | -0.03*** |
| Cohort3 | -0.06*** | -0.06*** | -0.06*** | -0.06*** |
| Cohort4 | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| Special education exit | 0.11* | 0.03 | 0.15 | -0.21 |
| $R^2$ | 0.0016 | 0.0016 | 0.0016 | 0.0016 |
| *Sample size (N)* | 342,211 | 344,211 | 341,262 | 341,288 |

*Note.* Cohort parameter estimates are calculated in relation to the fifth (2001-2003) cohort of students.
*Note.* A negative *B* for special education exit would represent positive special education gains.
* $p < .05$, ** $p < .01$, *** $p < .001$.

*Summary of Hypothesis Testing*

When considered together, these results reveal support for the hypotheses presented in Chapter Three. Below, Table 27 briefly summarizes the three hypotheses and the results found for each regression analysis. Because results for students entering special education programs are considered most indicative of the effectiveness of special education programs, only the parameter estimates (*B*) for special education entry and the overall $R^2$ of those models are presented for both reading and math assessments.

Table 27

*Summary of Hypothesis Testing*

| Hypothesis | Sp. ed. entry *B* Reading | Sp. ed. entry *B* Math | $R^2$ of model Reading | $R^2$ of model Math | Hypothesis supported? |
|---|---|---|---|---|---|
| *H1 (special education entry is a positive predictor of differenced gain scores)* | | | | | |
| All disabilities | .14*** | .11*** | .0011 | .0019 | Rdg, Math |
| SLD | .20*** | .15*** | .0012 | .0020 | Rdg, Math |
| SLI | -.02 | -.03 | .0008 | .0018 | No |
| BED | .15** | .08 | .0008 | .0018 | Rdg |
| OHI | .13*** | .09* | .0009 | .0018 | Rdg, Math |
| *H2 (special education entry is a positive predictor of differenced residual gain scores)* | | | | | |
| All disabilities | .19*** | .18*** | .0016 | .0020 | Rdg, Math |
| SLD | .25*** | .23*** | .0017 | .0021 | Rdg, Math |
| SLI | .00 | -.01 | .0010 | .0016 | No |
| BED | .15** | .16*** | .0010 | .0016 | Rdg, Math |
| OHI | .19*** | .16*** | .0012 | .0016 | Rdg, Math |
| *H3 (special education entry is a positive predictor of differenced residual gain scores after controlling for use of accommodations)* | | | | | |
| All disabilities | .14*** | .12*** | .0023 | .0030 | Rdg, Math |
| SLD | .19*** | .17*** | .0023 | .0029 | Rdg, Math |
| SLI | -.02 | -.02 | .0016 | .0022 | No |
| BED | .10* | .10* | .0016 | .0022 | Rdg, Math |
| OHI | .14*** | .11*** | .0017 | .0023 | Rdg, Math |
| *H3 (special education entry is a positive predictor of differenced residual gain scores after excluding students who received accommodations)* | | | | | |
| All disabilities | .04 | .05* | .0011 | .0016 | Math |
| SLD | .09** | .08* | .0011 | .0016 | Rdg, Math |
| SLI | -.03 | -.01 | .0011 | .0016 | No |
| BED | .07 | .18 | .0011 | .0016 | No |
| OHI | -.03 | -.02 | .0011 | .0016 | No |

* $p$ < .05, ** $p$ < .01, *** $p$ < .001

# CHAPTER SIX

## Discussion

For over 200 years, educators and researchers have voiced concerns regarding the quality of education provided to children with disabilities. Over the course of these 2 centuries, the focus of this concern has shifted from ensuring that children with disabilities receive access to a formal education to examining the effectiveness of special education programs for improving their academic achievement. Recently, provisions in the 1997 reauthorization of the Individuals with Disabilities Act and the 2001 authorization of the No Child Left Behind Act made improvement in the educational outcomes for children with disabilities a focus by requiring that (a) students with disabilities participate in large-scale assessments of academic achievement, and (b) states be held accountable for the achievement outcomes of all students, including those with disabilities. Now that students with disabilities are required to participate in large-scale assessments of academic achievement, researchers have a new tool for measuring academic outcomes for students with disabilities and examining the effectiveness of special education programs.

Hanushek et al. (1998) carried out one of the first studies to use large-scale achievement test data to assess the effectiveness of special education programs for children with disabilities. Their examination of students’ gain scores on a large-scale assessment of student achievement in Texas revealed that special education placement resulted in an overall increase in achievement scores for children with disabilities, with stronger effects reported in the area of math than in the area of reading. This research was important because it used state-level achievement data as a means to estimate the effectiveness of special education services. The current study aimed to replicate and expand upon Hanushek et al.'s research by using a similar methodology and extant data from North Carolina to estimate the effectiveness of special education programs. Similar to Hanushek et al.'s research, differences in yearly achievement gains on high-stakes assessments for students transitioning in and out of special education programs were examined. This research was intended to strengthen the external validity of Hanushek and colleagues' findings by applying similar methodology within another state. This study also aimed to extend this research by taking into consideration two additional factors that may have affected Hanushek et al.'s results: (a) regression to the mean and (b) students' use of accommodations during large-scale, high-stakes assessments.

The first five chapters reviewed previous research and legislation regarding the education of children with disabilities; described the three hypotheses for the current study; provided details regarding the methodology of the current research, including participants involved and measures utilized; and stated the results that were found after testing each hypothesis using multiple regression analyses. Chapter Six will begin with a discussion of these results, as they are related to the three study hypotheses, as well as previous research. Second, a general discussion will explore the significance of the current study in

the field of special education research. Then, the limitations of the current study and possible opportunities for future research endeavors will be presented. Finally, the broad implications of this research for special education practice will be discussed.

*Special Education Entry as a Significant Predictor of Achievement Growth*

The current study sought to examine the effectiveness of special education programs across the state of North Carolina by looking at how a select group of students with disabilities performed on large-scale achievement assessments. Five cohorts of students were followed as they progressed from the third through fifth grades. Students whose special education status changed during the fifth grade (i.e., they either entered or exited special education programs) were of particular interest. The achievement gains these students made while enrolled in general education programs were compared with the gains they made while enrolled in special education programs to estimate the effectiveness of special education on student achievement. Students' *z*-score score gains on End-of-Grade Tests of Reading Comprehension and Mathematics were expected to be greater when students participated in special education programs than general education programs. In accordance with the previous findings of Hanushek et al. (1998), these effects were expected to be more prominent for students who entered into special education programs than for students who exited from special education programs between the fourth and fifth grade. Three hypotheses were tested to determine if students' entrance or exit from special education programs was a significant predictor of their differing achievement gain scores on high-stakes assessments of reading and mathematics. Each hypothesis was confirmed, although the results were more consistently positive for the group of students entering special education programs compared to the group of students exiting special education programs. The bulk of this discussion will focus on the results for the group of students who entered special education programs, in keeping with the focus of Hanushek et al. They focused their discussion on this group because of concerns about the interpretability of results for students who exit special education programs due to the potential lingering effects of interventions even after a student stops receiving special education services.

*Hypothesis 1*

First, a replication of Hanushek and associates' (1998) key regression analyses was conducted. It was hypothesized that receipt of special education services would be a positive predictor of academic gain in both reading comprehension and mathematics for all students who experienced a change between general and special education programs. The effects for students who entered and exited special education programs were calculated together and separately.

Results supported the first hypothesis and were consistent with those found by Hanushek and colleagues (1998): after controlling for the effects of school-level characteristics and cohort, receipt of special education services was associated with a positive, significant gain in achievement in both reading and math. Similar to Hanushek et al., the effects of participation in a special education program were demonstrated more clearly by those students entering special education programs than those exiting special education programs. Entering special education programs was associated with a gain of 0.14 standard deviation in reading comprehension and a gain of 0.11 standard deviation in mathematics in *z*-score units when compared with students who remained in general education or special education programs over time. Taking into account the confidence limits around the parameter estimates obtained in the current study, these gains were larger in reading (0.14 versus 0.08) and smaller in mathematics (0.11 versus 0.17) than the gains reported by Hanushek et al.

For Hanushek et al. (1998), special education placement continued to be associated with a significant gain in achievement in reading and mathematics when students identified as SLD or BED were examined separately. In the current study, entry into special education programs was associated with significant gains in reading for students with SLD, BED, or OHI, and significant gains in mathematics solely for students with SLD or OHI. Similar to the findings of Hanushek et al., differences in achievement gains for students with SLI who were entering special education programs failed to reach significance in either academic area, and these students' gains were significantly smaller than the gains obtained for the other subgroups of students with disabilities (SLD, BED, and OHI). This pattern of results is reminiscent of previous findings by Carlberg and Kavale (1980) in their meta-analysis. They found that the effects of special education placement varied by subgroup of students with disabilities. Differential effects of special education by disability classification will be further addressed within the General Discussion section.

In sum, results from the set of analyses testing Hypothesis 1 strengthen the external validity of Hanushek and colleagues' (1998) research by replicating their findings using data from another state. Similar to their results, when students serve as their own controls and achievement gains made while in general education programs are compared to achievement gains made while in special education programs, entrance into special education is associated with gains in reading and mathematics.

*Hypothesis 2*

One of the two ways in which the current study aimed to expand upon the research by Hanushek et al. (1998) was to account for the effects of regression to the mean on students' achievement test score gain. Thus, Hypothesis 2 predicted that the advantages of special education over general education placement for students in the areas of reading comprehension and mathematics would continue to exist after controlling for regression to the mean.

This hypothesis was included based on a concern that the effects of regression to the mean would have biased the results of the current study in favor of the first hypothesis. As is described in Leary (2004), regression to the mean occurs because measurement error is correlated with obtained scores (but not with true scores). That is, individuals with obtained scores below the mean are more likely to have had their scores lowered due to "unlucky" measurement error and students with high scores are likely to have their obtained scores increased due to "lucky" measurement error. Upon retesting, the true score remains the same, but the testing error distributes itself randomly around its own mean (the standard error of measurement), resulting in regression of scores toward the mean.

Special education services are provided to a student whose disability affects his or her academic achievement in one or more areas (PSNC, 2007). Consequently, achievement scores for students with disabilities are likely to be below the mean, and this was true for participants within the current study (refer to Tables 6, 7, and 8, which report the third, fourth, and fifth grade *z*-scores for students with disabilities relative to the entire student population for that grade). Thus, for these students, gains between pre- and post-test scores may at least partially be due to the effects of regression to the mean and the previous results may overestimate the true gains made due to other changing factors (e.g., special education placement).

In testing the second hypothesis, regression to the mean was controlled for by predicting the difference in students' residual gain scores over the course of 3 years (i.e., the difference between the residual gains made during the fourth grade and the fifth grade), rather than their *z*-score gain over time. Therefore, the regression formula predicted the difference between students' fourth and fifth grade residual gain scores based on their changing special education status (i.e., exiting or entering special education programs) after controlling for the effects of school-level characteristics (i.e., the percent of students who received a free or reduced price lunch, the percent of students who were Black, and the percent of students who were Hispanic) and cohort.

Support for the second hypothesis was demonstrated in the areas of both reading and mathematics. Specifically, entering special education programs was associated with a gain of 0.19 standard deviation in reading comprehension and a gain of 0.18 standard deviation in mathematics when compared with students who remain in general education or special education programs over time. These results ruled out the possibility that previous results were solely a function of students' scores regressing toward the mean upon retesting.

It is interesting to note that for students who entered into special education programs during the fifth grade, the differenced residual gain scores were actually greater than the differenced gain scores without this correction (see Table 13). This finding was surprising because with special education students most likely to have scores lower than the mean, it seemed logical to predict that residual gain scores would decrease the amount of gain observed because a correction for regression toward the mean had been made.

A closer look at the mean *z*-scores for students who entered special education programs (see Table 6) is helpful in understanding why the differenced residual gain scores were greater than the differenced gain scores. Students who later entered special education programs demonstrated a mean *z*-score of -0.67 on the third grade reading test. These students' fourth grade reading scores were expected to regress toward the mean due to the redistribution of test error alone. However, the mean *z*-score in reading for these students actually decreased during the fourth grade to -0.79. With the mean score moving in the reverse direction predicted by regression toward the mean, the resulting residual gain score for these students was even more negative than their gain score. When these students took the test again in the fifth grade, their mean score increased. In fact, these students made much greater gains during the fifth grade than any other group of students. Although the fifth grade residual gain scores were smaller than the gain scores because of the adjustment for regression toward the mean, when fourth grade residual gain scores were subtracted from fifth grade residual gain scores, the resulting mean differenced residual gain scores were both positive and greater than the differenced gain scores.

Entry into special education programs continued to be a significant predictor of achievement gains in the areas of both reading and mathematics for students with SLD, BED, or OHI; once again, students with SLI did not demonstrate significant gains in either academic area. The effectiveness of special education programs beyond the effects of regression toward the mean continued to vary by disability type.

*Hypothesis 3*

The second way in which the current study aimed to expand upon the research by Hanushek et al., (1998) was to account for the effects of students using accommodations on large-scale achievement assessments, particularly when they are identified as having a disability. According to Hanushek et al., “despite controlling for all time-invariant unobserved differences among students, a potential problem with the fixed effect approach employed here is the assumption that changes in special education status are not accompanied by other changes that affect achievement” (p. 2). Although Hanushek and colleagues did address a variety of potential biases that may have affected their results, students’ use of test accommodations was not one of them. Therefore, in the current study, it was important to determine whether the results found after testing Hypotheses 1 and 2 would persist after controlling possible increases in achievement scores solely due to increased use of accommodations. The third hypothesis was tested using two different methods. First, students’ use of accommodations was entered as a separate independent variable into the prediction equation. Second, the regression analyses were run solely for students who did not receive accommodations on the EOG Tests during the third, fourth, and fifth grades.

*Controlling for accommodation use.* When students' use of accommodations was entered as a separate variable in the prediction equation, students' differing use of accommodations was a significant, positive predictor of their achievement gains. Thus, students' use of accommodations on large-scale achievement tests appears to account for part of the gain in achievement observed when students enter special education programs. However, special education entry itself also remained a significant predictor of students' differing residual gain scores in reading and math, providing support for the third hypothesis. After controlling for both regression to the mean and students' use of accommodations, entering special education programs was associated with a gain of 0.14 standard deviation in reading comprehension and a gain of 0.12 standard deviation in mathematics when compared with students who remained in general education or special education programs over time. These effect sizes were slightly lower than those found before controlling for accommodation use, suggesting that the use of testing accommodations is one of the benefits of receiving special education services. After controlling for both regression to the mean and students' use of accommodations, current estimates of special education effectiveness remained equal to or higher than those reported by Hanushek et al. (1998). These results ruled out the possibility that academic gains observed for students entering special education programs are solely due to the increased use of accommodations on large-scale assessments when students are enrolled in special education programs.

Results continued to vary by disability type. Entry into special education programs was a significant predictor of achievement gains in the area of reading and math for students with SLD, BED, or OHI. Again, differences in achievement gains for students with SLI who were entering special education programs failed to reach significance in either academic area.

*Excluding students who received accommodations.* The third hypothesis was also tested by simply excluding all students who received any accommodations during the 3 years of the study. Solely including students who never received any accommodations during the third, fourth, and fifth grade assessments decreased and changed the sample being assessed, such that over 98% of the sample consisted of students without disabilities and just over 1% of the sample consisted of students who entered or exited special education programs (compared to 89.5% of the sample consisting of students without disabilities, 1.7% of the sample consisting of students who entered special education programs and 1.3% of the sample consisting of students who exited special education programs in the primary analyses). Test accommodations are provided to students with disabilities in order to improve the validity of their test scores (Niebling & Elliott, 2005). Thus, it is assumed that this sub-sample included students with even milder disabilities, as their IEP teams did not determine that it was necessary for them to use accommodations on the EOG Tests.

When this sub-sample was considered, only partial support was found for the third hypothesis. Special education entry continued to remain a significant predictor of students' differing residual gain scores in mathematics: entering special education programs was associated with a gain of 0.05 standard deviation when compared with students who remained in general education or special education programs over time.

However, non-significant changes in residual gain scores were found in the area of reading. In terms of disability subgroups, for students with SLD, positive, significant results for entering special education programs continued to be found in both reading and math. Accommodations are typically a part of the special education placement package and are imperative in allowing students with disabilities to demonstrate their true knowledge on standardized tests (Koretz & Barton, 2003: Niebling & Elliott, 2005). Taken together, the current findings relative to Hypothesis 3 suggest that the achievement gains made by students with disabilities when they receive special education services may be partially due to their use of accommodations on these tests.

## *General Discussion*

This general discussion will explore the significance of the current study in the field of special education. Specifically, the contribution of the present study to the debate about the effectiveness of special education programs will be presented, as will possible explanations for the varying patterns of results observed within the two academic areas and for students with different types of disabilities. Finally, the evidence the current study provides for the positive impact of special education on achievement, even after controlling for the effects of access to test accommodations will be discussed.

### *Estimating the Effectiveness of Special Education Programs for Students with Disabilities*

The current study addressed the issue of whether special education programs are associated with improvements in academic outcomes by examining achievement gains on high-stakes assessments of reading and math when students transition between general and special education programs. The present finding that special education placement is associated with improvements in students' achievement scores is a relatively new finding in special education research.

With the exception of Hanushek et al. (1998), previous research has tended to reveal mixed results for special education placement. Notably, one of the most cited special education meta-analyses reported an overall effect size of -0.12 for special education placement (Carlberg & Kavale, 1980). Although the type of effect size computed by Carlberg and Kavale differs from the type of effect sizes calculated in the current study and by Hanushek et al. (1998), the positive findings in the more recent studies call into question earlier descriptions of special education placement as an intervention that does not work (Forness et al., 1997).

*Improving reading and math achievement.* Hanushek and colleagues' original 1998 findings suggested that special education entry was associated with larger achievement gains in the area of mathematics than in the area of reading. However, within the current study, special education entry was associated with gains in both academic areas (although there was a small, but significantly higher gain in the area of reading).

Hanushek et al. (1998) provided the following statement in a footnote to explain why stronger results were likely found in the area of math than reading:

> The larger impact of schools on math than reading for regular education could be explained by parents' having greater ability to help children in reading. It is less obvious that schools should have a larger impact in math for special education students, since anecdotal evidence suggests that reading problems are often

central to evaluation and classification for special education. Reading programs also appear plentiful in special education (p. 17).

This speculation seems limited in addressing why special education programs would be more effective in improving students' math achievement than their reading achievement. In fact, based on recent reading research published after Hanushek and colleagues' (1998) study, it is apparent that the knowledge base regarding effective reading instruction is greater than the knowledge base regarding effective mathematics instruction. Therefore, it seems likely that schools could be more effective in improving the reading achievement of students with disabilities than improving their math achievement.

Most students with learning disabilities (who comprise the largest group of students in special education) are reported to experience greater difficulties in the area of reading than in the area of math and, thus, more research has been conducted to understand the process of learning to read and how to intervene with students who have difficulties in the area of reading (Lyon et al., 2001). Research has indicated that effectively teaching children to read includes instruction in the following five areas: phonemic awareness, phonics, fluency, vocabulary, and text comprehension (National Institute for Literacy, 2001). Additionally, the use of strategies such as engaging in guided reading lessons, receiving instruction in word knowledge, writing, and independent reading have been demonstrated to improve the reading abilities of students with reading difficulties and disabilities (Richek, Caldwell, Jennings, & Lerner, 2002). Therefore, it is surprising that Hanushek et al. (1998) consistently found effects to be greater in the area of math than in the area of reading. Based on this recent surge in research and push to improve students' reading abilities, it is likely that reading instruction for students with disabilities has improved over the decade since Hanushek and colleagues' study. This change in instruction may explain why special education placement was associated with relatively equal gains in both reading and mathematics in present study.

*Improving achievement for students with different disabilities.* Although the results of this study indicated more favorable effects of special education when students with disabilities are considered as a group than have been demonstrated in previous research (e.g., Carlberg & Kavale, 1980), current results supported previous findings that the effectiveness of special education services varies for students with different types of disabilities.

For students with SLD, entering into special education programs resulted in higher gain scores in both reading (0.19) and math (0.17) than did participating in general education programs, even after controlling for the effects of regression to the mean and the use of accommodations. These results are similar to those of Carlberg and Kavale (1980), who found the effect size for special class placement for students with mild disabilities to be positive.

Similarly, students with other types of mild disabilities also experienced greater gains when placed in special education programs than when participating in general education programs. For students with BED, entering into special education programs was associated with higher gain scores in both reading (0.10) and math (0.10) than participating in general education programs, even after controlling for the effects of regression to the mean and the use of accommodations.  For students with OHI, entering into special education programs was associated with higher gain scores in both reading (0.14) and math (0.11) than participation in general education programs, even after controlling for the effects of regression to the mean and the use of accommodations.

Although support for the three hypotheses was consistently found in the areas of both math and reading and across disability types for entry into special education, results contrary to the hypotheses were occasionally found for students with specific disability types who exited special education programs.  Specifically, students with BED who exited special education programs demonstrated larger gains in the area of math when in general education programs, even after regression toward the mean and use of accommodations were controlled.  Students with OHI demonstrated a similar pattern in the area of reading.  Differences in the types of disabilities these students have, as well as the services they are receiving may account for some of these unexpected findings.

Students with BED or OHI include a very diverse group of students.  According to the National Association of School Psychologists (NASP; 2005), students receiving services for BED may have diagnoses such as schizophrenia, anxiety disorders, mood disorders, or other behavioral or emotional disturbances.  Similarly, students receiving special education services for OHI have chronic or acute health problems that affect their educational performance.  Students receiving services under the OHI category frequently have diagnoses such as heart conditions, chronic lung disease, tuberculosis, asthma, sickle cell anemia, epilepsy, leukemia, diabetes, or Attention Deficit/Hyperactivity Disorder (ADHD), among others (PSNC, 2007).

In addition to being diverse, the BED and OHI populations have also changed over time. Relatively recently, the number of students qualifying for special education students under the OHI disability category has grown at a significant rate (Schnoes, Reid, Wagner, & Marder, 2006). It was not until the 1997 reauthorization of the IDEA that federal policy explicitly included students with ADHD as eligible to be served through the OHI category. Therefore, it was not a frequently utilized classification category in which students with disabilities were placed to receive special education services at the time the data utilized by Hanushek et al. (1998) were collected. However, within the current sample, OHI was one of the four largest disability categories of students.

The increase in the number of students classified as OHI is attributed to increasing numbers of students with ADHD being shifted from receiving special education services under the BED label to receiving these services under the OHI label (Schnoes et al., 2006). However, prevalence rates of students with various disabilities (e.g., ADHD, mood disorders) in both the BED and OHI categories cannot be estimated within the current sample, as schools did not report this information. Due to their diverse and changing populations, it is somewhat difficult to make generalizations about individuals receiving special education services under the BED and OHI categories.

Because of the diverse behavioral, emotional, and/or medical difficulties that these students experience, the special education services that they receive are likely quite diverse as well. Students with BED as their primary disability likely receive special education services that focus on improving their emotional or behavioral difficulties, rather than solely and specifically targeting their academic skills in the areas of reading or math. Interventions often include individual and group counseling, social skills training, and career counseling or vocational planning, as well those related to improving their academic skills. In fact, the NASP stated that for students with BED, "academic difficulties often take a back seat to the student's behavioral difficulties" (p. 3). Thus, the services provided to students with BED or OHI (particularly those in the growing ADHD population) may influence the students' behavior more significantly than their academic achievement scores. Once these students are dismissed from special education services, the effects of the behavioral, emotional, and/or health interventions they have received may be more likely to continue to impact the students, even after they are participating in a purely "general education" curriculum. Therefore, the effectiveness of special education programs may be less apparent after they have been dismissed from special education services.

Likewise, students with speech-language impairments also represent a unique set of students with somewhat different intervention needs. Within both the study by Hanushek et al. (1998) and the current study, students with SLI demonstrated a different pattern of results than any other disability type. In both studies, entry into special education programs was not found to be a significant predictor of achievement gains for students with SLI. Perhaps the replication of this finding suggests that special education placement is not an effective intervention for improving the academic achievement of students with this type of disability. However, other possible explanations for this finding exist.

First, students with SLI as their primary disability are the least disabled students within the current sample (see Tables 10 and 11), as was true in the study by Hanushek et al. (1998). Students with other types of disabilities (e.g., learning disabilities, behavioral-emotional disabilities, mental disabilities) may or may not have a speech-language impairment as a secondary disability. However, students with SLI reported as their primary disability have not had any other disabilities identified within the school system. Within the current study (as well as Hanushek et al.'s study), students with SLI made up the largest percentage of students who exited special education programs during the fifth grade (45.7%). Much smaller percentages of students who entered special education programs during the fifth grade (11.2%) or remained in special education programs throughout the study (5.8%) had SLI as their primary disability. Additionally, students with SLI tended to have higher *z*-scores for their third, fourth, and fifth grade test scores than did students with SLD, BED, or OHI. Together, these statistics reveal that students with this type of disability have less severe impairments than other students within the current sample.

Second, most students with SLI as their primary disability likely receive special education services that focus on improving their overall communication skills, rather than specifically targeting their reading and math skills. Although the role of speech therapists has grown to include intervening with students in the areas of literacy development and social-emotional communication skills, their role has traditionally focused on providing interventions that improve students' oral language abilities, including articulation difficulties and receptive and expressive language skills (American Speech-Language-Hearing Association, 2000). Because they tend to have more mild disabilities and special education services often are directed at improving their oral communication skills, it is possible that students with SLI would not demonstrate significant improvements on large-scale assessments of reading comprehension and mathematics as a result of their special education placement. A more sensitive measure of oral language skills may better reflect the effectiveness of special education services for students with primary speech-language impairments.

In sum, the current study supports the previous finding by Carlberg and Kavale (1980) that the effectiveness of special class placement varies for students with different types of disabilities. Although special education placement is associated with improved achievement for most subgroups, the present study failed to find evidence of this benefit for students with speech-language impairments, and found some evidence that for students with behavioral-emotional disabilities and other health impairments, general education programs are associated with higher achievement gains.

*Improving achievement beyond the effects of accommodation use.* When students with disabilities participate in large-scale achievement tests, they often receive testing accommodations in accordance with their IEP. These accommodations are typically provided as one part of a special education program because they permit students with disabilities to demonstrate their true academic knowledge on achievement tests without being hindered by their disability (Koretz & Barton, 2003; Niebling & Elliott, 2005).

Previous research has demonstrated that students with disabilities tend to perform better on large-scale achievement tests when they are provided with accommodations than when they are not provided with accommodations; the scores of students without disabilities are not improved to the same degree when they are provided with similar accommodations (Elliott et al., 2001; Fletcher et al., 2006). Thus, it was important for the current study to estimate the effectiveness of special education programs beyond the impact of accommodation use.

Results supported the notion that accommodations are one of the benefits of receiving special education services. Receipt of accommodations was associated with a gain of 0.08 standard deviation in reading comprehension and a gain of 0.10 standard deviation in mathematics, independent from the effects of receiving special education services. Thus, simply receiving accommodations was associated with a significant gain in students' achievement scores. Additionally, when students' use of accommodations was controlled for, the effectiveness of special education entry decreased in both academic areas. When students who did not receive accommodations were considered separately, the achievement gain associated with special education entry decreased markedly in the area of math and was not different from zero in the area of reading. However, as noted earlier, eliminating students who had received accommodations meant only a select group of students were considered. Although both entry into special education programs and receipt of accommodations are independently associated with achievement gains, the results of the present study suggest that one advantage of special education placement is that students identified as having disabilities become eligible to receive accommodations on these large-scale assessments in accordance with their IEP.

## *Limitations*

As with all studies, the findings of the present study must be considered in light of the study's limitations. Six important limitations to the current study will be presented: flaws with the regression-based research design, limitations of using *z*-score gains, limitations of not controlling for students' school changes, questions concerning the representativeness of the sample, the effects of missing data on the results, and changes within the EOG Tests.

### *Research Design*

The first limitation of the current study is that it utilized a regression-based research design. No direct manipulation of variables occurred and no direct causal link between special education placement and achievement test score gain can be established. However, random assignment to special education and general education for students with disabilities presented legal, ethical, and logistical challenges that were beyond the scope of the present study.

Hanushek and colleagues' (1998) procedure of examining change in students' test score gain before and after entering or exiting special education placement is an innovative approach to estimating the effectiveness of special education, which does not require random assignment but circumvents some of the methodological problems with other approaches. Nonetheless, no definitive causal links between special education placement and achievement score gains can be drawn from the regression-based design of the current study.

*Use of Normalized Scores*

Second, the use of *z*-scores by year allows for the detection of gains for students in special education programs relative to students in general education programs; however, it does not allow for the detection of absolute gains in scores. If all students improved equally, *z*-score differences would remain the same, yet a mean score increase would be undetectable. The use of *z*-scores by year was chosen in order to most accurately replicate and expand upon the research of Hanushek et al. (1998). However, other methods for also considering mean changes in scores are available. For example, educators in the state of North Carolina use a growth formula that calculates *z*-scores based on the mean and standard deviation during a reference year. Therefore, both relative gains and gains made by the entire population are evident (PSNC, 2004c).

*Not Controlling for School Changes*

Third, within the present study, students' transfers between schools were not followed or controlled. School-level data (i.e., the percent of students who were Black, Hispanic, or receiving a free or reduced price lunch at a school) were taken from the students' fifth grade year and were not followed across the third through fifth grades. Therefore, if a student changed schools within the state over the course of these 3 years, their school-level information was not altered. Additionally, students' changes between schools were not controlled for within the regression equation. In their 1998 research, Hanushek and colleagues ran their analyses separately for all students, as well as those who did not change schools during the course of the study and their results did not change. Therefore, within the current study, student's school changes were not controlled. However, as school characteristics were consistently associated with differenced achievement score gains, it is possible that controlling for students' school changes would have minimally affected the results.

*Sample Representativeness*

Fourth, the representativeness of the current sample relative to the overall population of special education students is unclear. The current study only truly estimates the effectiveness of special education programs for "high functioning" students with disabilities that were likely mild enough that special education placement was delayed until the fifth grade, when they first began receiving special education services. It is possible that these students are more likely to respond positively to special education interventions, as well as to demonstrate their knowledge more easily on achievement tests, than are students who began receiving special education services at a younger age. Similarly, it has been suggested that higher achieving students with disabilities are likely to be dismissed from special education programs and it is expected that the same pattern occurred in the current sample (Ysseldyke & Bielinski, 2002). Together, students who entered or exited special education programs during the fifth grade solely constituted 28% of the current sample of students with disabilities. Because students who enter or exit special education programs represent a smaller, higher-functioning subset of the total population of students with disabilities, the effectiveness of special education programs may have been somewhat overestimated and these results may or may not generalize well to students with more severe disabilities (e.g., cognitive disabilities).

Additionally, not all students with disabilities participated in the EOG Tests each year. Although the participation rates by students with disabilities was found to be much higher in the current study than at the time of Hanushek et al.'s (1998) research, current results cannot necessarily be generalized to the students who were exempted from the EOG Tests.

*Missing Data*

There are data that were unavailable during the current research, but that would have been beneficial to the study. The first is information regarding the organizational or instructional practices that characterized the special education services students received. It would have been interesting to determine if specific classroom variables, such as the pupil-teacher ratio, the qualifications of the teacher, and the types of special education services provided (e.g., inclusion within the general education classroom, pull-out resource services, instruction within self-contained classrooms) were associated with greater achievement gains for students with different types of disabilities. Additionally, instructional variables (i.e., describing what "special education" involves) would also have been interesting to consider. As was noted in Chapter Two, the definition of a "special education" treatment is a continuing question in special education research (e.g., Fuchs & Fuchs, 1995; Hocutt, 1996; Scruggs & Mastropieri, 1995), as well as in all treatment research. Data regarding the classroom and instructional variables that constituted special versus general education programs were unavailable during the current study.

One particular instructional variable that would be of primary interest is the length of a student's special education intervention. Specifically, it would have been helpful to have a record of when students' special education status changed during their fifth grade year. All that can be determined from the current database is that students' disability status (either having a disability or not having a disability) changed from fourth to fifth grade. It is impossible to know the exact date when these students entered or exited special education programs. For example, if a student initially entered into a special education program at the beginning of the fifth grade, he or she would be exposed to special education interventions throughout his or her entire fifth grade school year before taking the EOG Tests in May. On the other hand, if a student initially entered into a special education program in the middle to end of the fifth grade, he or she would only be exposed to special education interventions for a few months (or possibly even weeks) before taking the EOG Tests in May. However, both of these students would be treated as students who entered into special education programs during the fifth grade and would be given the same disability status. The same is true for students who exited from special education services: some may have been dismissed at the beginning of the fifth grade and others may have been dismissed later during the school year. Thus, some students entering special education programs likely received less than 1 year of special education instruction and some students exiting special education programs likely spent less than 1 year involved only in a general education program; however, fifth grade achievement gain was interpreted for all students as 1 year of special versus general education placement and, therefore, 1 year of special versus general education achievement gain.

Second, students' disability classification (e.g., Specific Learning Disability, Other Health Impaired) was solely determined at the school level; that is, there was no external verification that students met criteria for any particular disability classifications. Although the procedures for identifying students with disabilities across the state of North Carolina are specified in the *Policies Governing Services for Children with Disabilities* (PSNC, 2007), different school districts, schools, and even individual school psychologists and IEP teams may have used different decisions for classifying students with different types of disabilities. This concern is particularly true of the Specific Learning Disability category, which has historically been difficult to define. In 1983, Shepard, Smith, and Vojir revealed that less than 50% of a selected sample of Colorado students with SLD demonstrated characteristics that were consistent with federal definitions of the disability category. Concerns about the identification of students with SLD continue, as the diagnosis of this type of disability is currently changing from one based on an ability-achievement discrepancy to that of a response-to-intervention method (Holdnack & Weiss, 2006). Without external verification of students' disability status, the current study relied solely on the report of each school to determine each student's disability and special education placement. Thus, it is difficult to predict the generalizability of these results to different groups of students with disabilities. This difficulty is exacerbated by the fact that individual states use different criteria for special education disability classifications, as is permitted in federal legislation.

Also, it would be beneficial to have data regarding students' secondary disability classifications. In the current database, only primary disability classifications were provided. It is unknown whether students also have secondary disabilities, particularly speech-language impairments. It would have been interesting to examine the effectiveness of special education for students with single versus multiple disabilities as well.

The lack of information about programmatic or instructional variables, which are likely to account for gains children make in any educational setting, may partially explain why the control variables and special education transitions together accounted for such a small amount of the variance in students' differenced achievement score gains. In all of the regression analyses completed, the $R^2$ values were quite small (< 1%). By Cohen's (1988) criteria, these values fall well below his suggested value for even a "small" effect size in a regression-based design.

*Test Changes*

Finally, it is important to note that the EOG-Math Test changed during the 2000-2001 school year and the EOG-Reading Test changed during the 2002-2003 school year. The content of both tests was updated to reflect changes to the state's mathematics and language arts curricula respectively. With new editions, the tests' vertical scaling was lost so that scores for the new editions were not directly comparable to the previous editions. The present study did not use the vertically scaled test scores, but instead converted all scores to grade and year-based z-scores, in part to accommodate different test editions. However, it is possible that this change may have had some unknown effect on the score distributions for general and special education students. For example, Koretz (1996) reported that some of the gains observed across years on achievement tests are specific to the items on a test, rather than generalizable academic skills, so that when a different test or edition is used, test scores decrease. In the present study, there were cohort effects by year that were entered as controls in the regression equations. Some of these cohort effects may have reflected the changes in test editions, although the direction of the effect was not to lower scores across years. Instead, gains in the last year of the study were higher than those in previous years, as indicated by the negative beta weights found for the first four cohorts compared to the last cohort in the regression analyses.

## *Directions for Future Research*

The current study provides support for the effectiveness of special education programs by utilizing a new methodology that examines students' gain scores on large-scale assessments. It is important for future research to be continued in this area. Directions for future research are suggested based on the limitations of the current study, as well as other potential areas of interest.

### *Research Design*

First, the use of a hierarchical linear modeling approach would allow researchers to examine the effectiveness of special education programs while also taking into consideration the structure in which students' educational programming exists. Students with and without disabilities are almost always participants within a hierarchical structure that involves being nested within classes, within schools, within districts, within states. As Raudenbush and Bryk (2002) state, "educational research is often especially challenging because studies of student growth often involve a doubly nested structure of repeated observations within individuals, who are in turn nested within organizational settings" (p. 4). Using a hierarchical linear model would take into account differences between classrooms, schools, districts, and even states that affect student outcomes.

Second, within the current study, no direct manipulation of variables occurred and no direct causal link between special education placement and achievement test score gain can be established. Additionally, no data were available to examine the instructional variables that were used during the special education interventions students received. Hocutt (1996) suggested that it is the intervention provided, rather than the placement, per se, which improves outcomes for students with disabilities. Similarly, Scruggs and Mastropieri (1995) recommended that special education programs be evaluated based on their utilization of the PASS variables previously described (e.g., prioritizing objectives within the curriculum; adapting materials and instructional strategies to meet individual needs; ensuring that teachers' instructional presentation is clear, structured, enthusiastic, and engaging; and systematically monitoring students' progress). Hence, future research ought to estimate the effectiveness of special education interventions after directly manipulating important classroom and/or instructional variables (e.g., pupil-teacher ratio, teacher qualifications, type of instruction used, length of the intervention) and documenting the treatment integrity of the special education instruction. The current study demonstrated that special education programs are associated with improving the achievement of students with disabilities (particularly those with mild disabilities). What is it about special education interventions that is most beneficial for improving long-term student outcomes and preparing these students for high school graduation, post-secondary education, and/or the work force? Manipulation of the instructional variables utilized during special education instruction might allow for a greater amount of variance to be accounted for in students' achievement score gains.

*Further Examination of Data*

Finally, future research could involve simply examining the current dataset or other similar datasets further. For example, it would be interesting to examine the current sample more closely to examine which types of testing accommodations were most effective in improving the achievement scores of students with different disabilities in each subject area. Additionally, it might be interesting to consider the effectiveness of special education programs for students with multiple versus single disabilities. One area that might be of particular interest would be to follow up on the outcomes for students with speech-language impairments, who did not appear to benefit from special education services within the current study. The effectiveness of special education programs for these students may be more easily demonstrated through students' gains on other educational outcome measures (e.g., standardized or curriculum-based assessments of articulation skills, language usage, and/or early literacy skills).

Conducting similar research in another state (perhaps one in yet another geographic location) would continue to improve the external validity of both Hanushek et al.'s (1998) study and the current study. Because different states use different disability criteria, the findings in the states of Texas and North Carolina may not necessarily be directly applicable to students with disabilities across the U.S. Additionally, with the passage of the NCLB Act, more students are likely to be included in large-scale assessments over time. Therefore, by replicating the current study a few years in the future, or within another state, a broader sample of students with disabilities may be available.

## *Implications for Practice*

The effectiveness of special education programs is an area of great concern to educators, researchers, and parents. In this time of increased accountability for individual schools, school districts, and states, the present study provides some support for the present configuration of special education programs in the education of students with disabilities, particularly those with more mild disabilities, such as specific learning disabilities or other health impairments. The results of the current study have direct implications for practice in terms of closing the achievement gap, funding special education programming, and utilizing evidence-based interventions.

### *Closing the Achievement Gap*

With the passage of the IDEIA and the NCLB Act, increasingly large numbers of students with disabilities are participating in high-stakes, large-scale assessments of achievement (Elliott et al., 2000). In addition to ensuring that students with disabilities are exposed to the general curriculum and are participating in state-wide assessments of academic achievement, these legislative acts also require that schools provide services that allow students with disabilities to reach 100% proficiency as a subgroup within the next 5 years, or face sanctions. Data presented by the American Youth Policy Forum and Center on Education Policy (2001) suggest that the majority of students with disabilities have demonstrated stability or growth on state-wide tests of academic achievement in recent years: “According to the annual surveys of the National Center on Educational Outcomes, the performance of students with disabilities on state assessments has increased in 28% of the states, remained about the same in 36% of the states, and decreased in 8% of the states… 28% of the states do not yet have data comparing their performance with previous years” (p. 28). Results of the current study suggest that providing special education services can help improve the achievement of students with disabilities. But is the achievement gap between students with and without disabilities in the state of North Carolina likely to close in the near future given current special education programming?

According to the results of the current study, after controlling for regression to the mean, entering special education programs was associated with a gain of 0.19 standard deviation in reading comprehension and a gain of 0.18 standard deviation in mathematics. As is demonstrated in Tables 6, 7, and 8, students with disabilities tend to have *z*-scores that fall between one-third to nearly one whole standard deviation below the mean in grades 3, 4, and 5. Optimistically assuming that the rate of achievement gain observed in the current study would continue each year a student received special education services, it would take a student with a disability approximately 3 to 5 years to improve his or her reading and math achievement scores by 0.5 to 1.0 standard deviation. Although there is no guarantee that these students will be performing at a level of *proficiency* by the year 2014, preliminary data suggest that the achievement gap has the potential to be narrowed considerably by providing special education services to students with disabilities.

In sum, how one interprets the present results in terms of the effectiveness of special education depends on the metric. In terms of total variance in achievement gain accounted for, the impact of special education is quite small (Cohen, 1988). However, when the results of current study are considered in a meaningful context, as recommended by Bloom, Hill, Black, and Lipsey (2008), the gains in achievement associated with special education placement in the current study are relatively positive. Special education placement is associated with a narrowing of the achievement gap between students with and without disabilities.

*Funding Special Education Programming*

It is well documented that special education programs designed to educate students with disabilities are expensive to implement (e.g., Chaikind et al., 1993; Chambers et al., 2002). In fact, one of the reasons Hanushek et al. (1998) aimed to estimate the effectiveness of special education programs was to provide additional validation for the "disproportionate amount of school funding that goes to the education of disabled children – perhaps as much as one-fifth of total current spending for slightly more than 10 percent of students" (p. 1). Hanushek and colleagues' results revealed that special education programs are associated with academic gains for students with disabilities; however, their results did not allow for a determination as to whether the benefits of special education interventions are large enough to justify the extra costs of educating a student with a disability. The same is true for the current study. Nevertheless, multiple studies have now reported on the positive effects of special education on the achievement of students with disabilities. Therefore, it seems appropriate to conclude that special education programming is a valuable method for educating students with disabilities that ought to continue to be funded.

*Utilizing Evidence-Based Interventions*

Educators should feel comforted by the knowledge that special education programming can advance the achievement of students with disabilities, particularly as schools, school districts, and states continue to feel the ramifications of recent legislation requiring that students with disabilities participate in and perform proficiently on large-scale assessments of their achievement. Because of its complexity, the validation of special education as a broad policy is not possible with evidence-based practice criteria such as those put forth put forth by Chambless, Baker, Baucom, Beutler, Calhoun, Crits-Christoph et al. (1998) or the U.S. Department of Education (2003). However, educators, administrators, and school psychologists can work to ensure that the specific interventions used in special education instruction are evidence-based in accordance with stipulations put forth by the NCLB Act (2001). According to the U.S. Department of Education, interventions such as one-on-one tutoring, reduction of class size in the early grades, and early reading instruction in phonemic awareness and phonics have demonstrated effectiveness during randomized controlled trials. As components of special education programs, utilization of these interventions can only help to increase the effectiveness of special education programs targeted at improving achievement for students with disabilities.

In summary, the present study aimed to estimate the impact of special education programs on achievement in the state of North Carolina by examining differences in the gain scores made by students with disabilities on large-scale achievement tests when they were enrolled in general versus special education programs. Results suggest that there are positive effects of special education on the reading and math achievement of students with disabilities. The current study replicates and expands upon previous research and suggests that the positive effects of special education placement are robust, and remain even after controlling for the effects of regression to the mean and students' use of accommodations on large-scale assessments. Although special education effectiveness varies for students with differing disabilities, students with mild disabilities (particularly those with specific learning disabilities or other health impairments) tend to benefit from participation in special education programs; the academic benefits for students with speech-language impairments are less clear at the present time. The current study adds additional evidence to the accumulating body of research indicating that students' scores on large-scale assessments are useful in providing information about the effectiveness of special education programs. Research utilizing the current methodology for assessing the effectiveness of special education programs has now been applied in at least two states and has revealed consistent, positive effects for the placement of students with disabilities in special education programs.

# REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington DC: Authors.

American Youth Policy Forum & Center on Education Policy. (2001). *Twenty-five years of educating children with disabilities: The good news and the work ahead.* Washington DC: Authors. Retrieved July 4, 2009 from www.aypf.org

American Speech-Language-Hearing Association. (2000). *Guidelines for the Roles and Responsibilities of the School-Based Speech-Language Pathologist* [Guidelines]. Retrieved March 4, 2009 from www.asha.org/policy

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.

Braden, J. P. (2002). Best practices for school psychologists in educational accountability: High stakes testing and educational reform. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 301-320). Bethesda, MD: National Association of School Psychologists.

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

Chaikind, S., Danielson, L. C., & Brauen, M. L. (1993). What do we know about the costs of special education? A selected review. *The Journal of Special Education, 26,* 344-370.

Chambers, J. G., Parrish, T. B., & Harr, J. J. (2002). *What are we spending on special education services in the United States, 1999-2000?* (Special Education Expenditure Project Report #1). Palo Alto, CA: American Institutes for Research, U.S. Department of Education, Office of Special Education Programs. Retrieved April 20, 2007 from http://www.csef-air.org/

Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., Crits-Christoph, P. et al. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist, 51,* 3-16.

Christenson, S. L., Decker, D. M., Triezenberg, H. L., Ysseldyke, J. E., & Reschly, A. (2007). Consequences of high-stakes assessment for students with and without disabilities. *Educational Policy, 21,* 662-690.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Crawford, L., Almond, P., Tindal, G., & Hollenbeck, K. (2002). Teacher perspectives on inclusion of students with disabilities in high-stakes assessments. *Special Services in the Schools, 18,* 95-118.

Deno, E. (1970). Special education as developmental capital. *Exceptional Children, 37,* 229-237.

Dorn, S., Fuchs, D., & Fuchs, L. S. (1996). A historical perspective on special education reform. *Theory Into Practice, 35,* 12-19.

Dunn, L. M. (1968). Special education for the mildly retarded: Is much of it justifiable? *Exceptional Children, 34,* 5-22.

Elliott, J. L., Erickson, R. N., Thurlow, M. L., & Shriner, J. G. (2000). State-level accountability for the performance of students with disabilities: Five years of change? *The Journal of Special Education, 34,* 39-47.

Elliott, S. N., Kratochwill, T. R., & McKevitt, B. C. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. *Journal of School Psychology, 39,* 3-24.

Fletcher, J. M., Francis, D. J., Boudousquie, A., Copeland, K., Young, V., Kalinowski, S., et al. (2006). Effects of accommodations on high-stakes testing for students with reading disabilities. *Exceptional Children, 72,* 136-150.

Forness, S. R., Kavale, K. A., Blum, I. M, & Lloyd, J. W. (1997). Mega-analysis of meta-analyses: What works in special education and related services. *Teaching Exceptional Children, 29,* 4-9.

Fuchs, D., & Fuchs, L. S. (1994). Inclusive schools movement and the radicalization of special education reform. *Exceptional Children, 60,* 294-309.

Fuchs, D., & Fuchs, L. S. (1995). What's "special" about special education? *Phi Delta Kappan, 76,* 522-530.

Gartner, A., & Lipsky, D. K. (1987). Beyond special education: Toward a quality system for all students. *Harvard Educational Review, 57,* 367-390.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*, 149-164.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (1998). *Does special education raise academic achievement for students with disabilities?* (Working Paper No. 6690). Cambridge, MA: National Bureau of Economic Research. Retrieved March 25, 2003, from http://www.nber.org/papers/w6690

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). Inferring program effects for special populations: Does special education raise achievement for students with disabilities? *The Review of Economics and Statistics*, *84*, 584-599.

Hocutt, A. M. (1996). Effectiveness of special education: Is placement the critical factor? *The Future of Children, 6*, 77-102.

Holdnack, J. A., & Weiss, L. G. (2006). IDEA 2004: Anticipated implications for clinical practice – Integrating assessment and intervention. *Psychology in the Schools, 43*, 871-882.

Huefner, D. (2000). *Getting comfortable with special education law*. Norwood, MA: Christopher-Gordon.

Johnson, E. S. (2000). The effects of accommodations on performance assessments. *Remedial and Special Education, 21*, 261-267.

Jones, B. D. (2007). The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology*, *23*, 65-86.

Kavale, K. A., & Forness, S. R. (1999). Effectiveness of special education. In C. R. Reynolds and T. R. Gutkin (Eds.), *Handbook of school psychology* (3rd ed., pp. 984-1024). New York: Wiley.

Kavale, K. A., & Forness, S. R. (2000). History, rhetoric, and reality: Analysis of the inclusion debate. *Remedial and Special Education, 21*, 279-296.

Koretz, D. (1996). Using student assessments for educational accountability. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (pp. 171-195). Washington, DC: National Academy Press.

Koretz, D. M., & Barton, K. (2003). *Assessing students with disabilities: Issues and evidence* (CSE Tech. Rep. No. 587). Los Angeles: University of California, Graduate School of Education and Information Studies, National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation. Retrieved March 21, 2005, from http://www.creest.org/products/reports_set.htm

Lang, S. C., Kumke, P. J., Ray, C. E., Cowell, E. L., Elliott, S. N., Kratochwill, T. R., et al. (2005). Consequences of using testing accommodations: Student, teacher, and parent perceptions of and reactions to testing accommodations. *Assessment for Effective Intervention, 31,* 49-62.

Leary, M. R. (2004). *Introduction to behavioral research methods* (4th ed.). Boston: Pearson Education.

Leinhardt, G., & Pallay, A. (1982). Restrictive educational settings: Exile or haven? *Review of Educational Research, 52,* 557-578.

Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B. et al. (2001). Rethinking learning disabilities. In C. E. Finn, A. J. Rotherham, C. R. Hokanson (Eds.), *Rethinking special education for a new century* (pp. 259-287). Washington, DC: Thomas B. Fordham Foundation and Progressive Policy Institute.

Madden, N. A., & Slavin, R. E. (1983). Mainstreaming students with mild handicaps: Academic and social outcomes. *Review of Educational Research, 53,* 519-569.

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.

McLaughlin, M. J., & Warren, S. H. (1992). Outcomes assessment for students with disabilities: Will it be accountability or continued failure? *Preventing School Failure, 36,* 29-33.

McLeskey, J. (2004). Classic articles in special education: Articles that shaped the field, 1960 to 1996. *Remedial and Special Education, 25,* 79-87.

Mehrens, W. A. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives,* [On-line serial], *6* (13). Retrieved March 25, 2001 from http://epaa.asu.edu/v6n13.html

National Association of School Psychologists. (2005). *Position statement on students with emotional and behavioral disorders.* Retrieved March 4, 2009 from http://www.nasponline.org/about_nasp/pospaper_sebd.aspx

National Center for Educational Statistics. (2006). *Digest of educational statistics, 2005.* Washington DC: U.S. Department of Education. Retrieved April 12, 2007 from http://nces.ed.gov/fastfacts/

National Institute for Literacy. (2001). *Put reading first: The research building blocks for teaching children to read*. Washington DC: US Government Printing Office.

Nichols, S. L. (2007). High-stakes testing: Does it increase achievement? *Journal of Applied School Psychology*, *23,* 47-64.

Niebling, B. C., & Elliott, S. N. (2005). Testing accommodations and inclusive assessment practices. *Assessment for Effective Intervention, 31*, 1-6.

No Child Left Behind Act of 2001, Reauthorization of the Elementary and Secondary Education act, Public Law No. 107-110 (2001).

Osborne, A. G., & Dimattia, P. (1994). The IDEA's least restrictive environment mandate: Legal implications. *Exceptional Children, 61,* 6-14.

Princeton Review. (2003). *Testing the testers 2003: An annual ranking of state accountability systems.* Princeton, NJ: Author.

Public Schools of North Carolina. (1998). *Interpretive guide for North Carolina End-of-Grade tests: Math, reading comprehension* (Local Use Form 1997-1998). Raleigh, NC: Public Schools of North Carolina, Department of Public Instruction, Raleigh Division of Accountability/Testing.

Public Schools of North Carolina. (2004a). *Understanding North Carolina end-of-grade testing* (Assessment Brief). Raleigh, NC: Public Schools of North Carolina, North Carolina Department of Public Instruction, Division of Accountability Services. Retrieved March 1, 2007 from http://www.ncpublicschools.org/accountability/testing

Public Schools of North Carolina. (2004b). *North Carolina reading comprehension tests: Technical report (Citable draft).* Raleigh, NC: Public Schools of North Carolina, North Carolina Department of Public Instruction, Office of Curriculum and School Reform Services. Retrieved February 26, 2006 from http://www.ncpublicschools.org/docs/accountability/testing/readingtechmanual.pdf

Public Schools of North Carolina. (2004c). *Setting annual growth standards: "The formula"* (Accountability Brief). Raleigh, NC: Public Schools of North Carolina, North Carolina Department of Public Instruction, Division of Accountability Services.

Public Schools of North Carolina. (2006). *The North Carolina math tests: Technical report.* Raleigh, NC: Public Schools of North Carolina, North Carolina Department of Public Instruction, Office of Curriculum and School Reform Services. Retrieved March 19, 2007 from http://www.ncpublicschools.org/docs/accountability/testing/mathtechmanual.pdf

Public Schools of North Carolina. (2007). *Policies governing services for children with disabilities.* Raleigh, NC: Public Schools of North Carolina, North Carolina Department of Public Instruction, Exceptional Children Division. Retrieved April 1, 2009 from http://www.ncpublicschools.org/ec/policy

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

Richek, M. A., Caldwell, J. S., Jennings, J. H., & Lerner, J. W. (2002). *Reading problems: Assessment and teaching strategies* (4th ed.). Boston, MA: Allyn & Bacon.

Roach, A. T., & Frank, J. L. (2007). Large-scale assessment, rationality, and scientific mangaement: The case of No Child Left Behind. *Journal of Applied School Psychology, 23*, 7-25.

Reynolds, M., Wang, M., & Walberg, H. (1987). The necessary restructuring of special education. *Exceptional Children, 53,* 391-398.

Schnoes, C., Reid, R., Wagner, M., & Marder, C. (2006). ADHD among students receiving special education services: A national survey. *Exceptional Children, 72,* 483-496.

Schulte, A. C. (1996). Remediation and inclusion: Can we have it all? In D. L. Speece & B. K. Kough (Eds.), *Research on classroom ecologies: Implications for inclusion of children with learning disabilities* (pp. 203-210). Hillsdale, NJ: Lawrence Erlbaum.

Schulte, A. C., Osborne, S. S., & Erchul, W. P. (1998). Effective special education: A United States dilemma. *School Psychology Review, 27,* 66-76.

Schulte, A. C., & Villwock, D. N. (2004). Using high-stakes tests to derive school-level measures of special education efficacy. *Exceptionality, 12,* 107-127.

Schulte, A. C., Villwock, D. N., Whichard, S. M., & 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.

Scruggs, T. E., & Mastropierei, M. A. (1995). What makes special education special? Evaluating inclusion programs with the PASS variables. *The Journal of Special Education, 29,* 224-233.

Shepard, L. A., Smith, M. L., & Vojir, C. P. (1983). Characteristics of pupils identified as learning disabled. *American Educational Research Journal, 20,* 309-331.

Shinn, M. R. (1986). Does anyone care what happens after the refer-test-place sequence: The systematic evaluation of special education program effectiveness. *School Psychology Review, 15,* 49-58.

Shriner, J. G., & DeStefano, L. (2001). Participation in statewide assessments: Views of district-level personnel. *Assessment for Effective Intervention, 26,* 9-16.

Sindelar, P. T., & Deno, S. L. (1978). The effectiveness of resource programming. *The Journal of Special Education, 12,* 17-28.

Stainback, W., & Stainback, S. (1984). A rationale for the merger of special education and regular education. *Exceptional Children, 51,* 102-111.

Texas Education Agency. (2003). *Timeline of testing in Texas.* Retrieved March 14, 2007, from http://www.tea.state.tx.us/student.assessment/resources/studies/testingtimeline.pdf

Thurlow, M. L., House, A. L., Scott, D. L., & Ysseldyke, J. E. (2000). Students with disabilities in large-scale assessments: State participation and accommodation practices. *The Journal of Special Education, 34,* 154-163.

U.S. Department of Education. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Washington DC: Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Will, M. C. (1986). Educating children with learning problems: A shared responsibility. *Exceptional Children, 52,* 411-416.

Wooldridge, J. M. (2000). *Introductory econometrics: A modern approach.* United States: South-Western College Publishing.

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.

Ysseldyke, J., Nelson, J. R., Christenson, S., Johnson, D. R., Dennison, A., Triezenberg, H., et al. (2004). What we know and need to know about the consequences of high-stakes testing for students with disabilities. *Exceptional Children, 71,* 75-94.

Zigmond, N. (2003). Where should students with disabilities receive special education services? Is one place better than another? *The Journal of Special Education, 37,* 193-199.

Zigmond, N., & Baker, J. M. (1996). Full inclusion for students with learning disabilities: Too much of a good thing? *Theory Into Practice, 35*, 26-34.

Zlatos, B. (1994). Don't test, don't tell: Is 'academic red-shirting' skewing the way we rank our schools? *The American School Board Journal, 181,* 24-28.