

Georgia State University

ScholarWorks @ Georgia State University

Communication Sciences and Disorders
Dissertations

Department of Communication Sciences and
Disorders

Spring 5-17-2013

Investigating the Effects of a Read-aloud Alteration on the Third-grade Reading Criterion-referenced Competency Test (CRCT) for Students with Disabilities

Melissa Fincher
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/epse_diss

Recommended Citation

Fincher, Melissa, "Investigating the Effects of a Read-aloud Alteration on the Third-grade Reading Criterion-referenced Competency Test (CRCT) for Students with Disabilities." Dissertation, Georgia State University, 2013.

https://scholarworks.gsu.edu/epse_diss/87

This Dissertation is brought to you for free and open access by the Department of Communication Sciences and Disorders at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Communication Sciences and Disorders Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

ACCEPTANCE

This dissertation, INVESTIGATING THE EFFECTS OF A READ-ALoud ALTERATION ON THE THIRD-GRADE READING CRITERION-REFERENCED COMPETENCY TEST (CRCT) FOR STUDENTS WITH DISABILITIES, by MELISSA FINCHER, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Education, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chair, as representatives of the faculty, certify that this dissertation has met all the standards of excellence and scholarship as determined by the faculty. The Dean of the College of Education concurs.

Nannette Commander, Ph.D.
Committee Chair

Daphne Greenberg, Ph.D.
Committee Member

Claudia Flowers, Ph.D.
Committee Member

Andrew Roach, Ph.D.
Committee Member

Date

Laura Fredrick, Ph.D.
Chair, Department of Educational Psychology and Special Education

Paul A. Alberto, Ph.D.
Interim Dean
College of Education

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulation governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education's director of graduate studies and research, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my permission

Melissa Fincher

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Melissa Fincher
1124 DeKalb Avenue, #5
Atlanta, GA 30307

The director of this dissertation is:

Dr. Nannette Commander
Department of Educational Psychology and Special Education
College of Education
Georgia State University
Atlanta, GA 30302-3979

CURRICULUM VITAE

Melissa Lane Fincher

ADDRESS: 1124 DeKalb Avenue, #5
Atlanta, GA 30307

EDUCATION:

2013	Ph.D.	Georgia State University Educational Psychology
1991	M.S.	Georgia State University Educational Research
1986	B.A.	University of Georgia Psychology

PROFESSIONAL EXPERIENCE:

2009-Present	Associate Superintendent for Assessment & Accountability Georgia Department of Education
2008-2009	Director of Assessment Georgia Department of Education
2007-2008	Director of Assessment Research and Development Georgia Department of Education
2004-2007	Assistant Director of Testing Georgia Department of Education
2002-2004	Research Associate Atlanta Public Schools
1997-2002	Educational Assessment Specialist Georgia Department of Education

PUBLICATIONS AND PRESENTATIONS

- Perie M., Fincher, M., and Swaffield, S. (2013). Understanding low-performing students with disabilities and their barriers to success on traditional Assessments: A southern tale. In Thurlow, M. L., Lazarus, S. S., & Bechar, S. (Eds.). *Lessons Learned in Federally Funded Projects that can Improve the Instruction and Assessment of Low Performing Students with Disabilities*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Fincher, M. (February 2012). *Detection and Analysis of Irregularities in Academic Testing: Georgia's Experience*. Testing Integrity Symposium: Issues and Recommendations for Best Practices, National Center for Educational Statistics, Washington, DC.
- Englehard, G., Fincher, M., & Domaleski, C.S. (2011). Mathematics performance of students with and without disabilities under accommodated conditions using resource guides and calculators on high stakes tests. *Applied Measurement in Education*, 24(1), 22 – 38.
- Roach, A., Fincher, M., Varjas, K., & Chilungu, N. (April 2011). *Exploring the Thought*

- Processes Underlying the Scoring of Georgia's Alternate Assessment*. National Council on Measurement in Education Annual Conference, New Orleans, LA.
- Fincher, M. & Flowers, C. (2009). Georgia alternate assessment: A portfolio approach. In W. D. Schafer and R. Lissitz (Eds.), *Technical Quality of Alternate Assessments based on Alternate Achievement Standards* (pp. 171-188). Baltimore: Brookes Publishing.
- Dunn, J. L., & Fincher, M. (March 2009). *A Framework for Identifying and Evaluating Item Alterations Designed for Persistently Low Performing Students*. Paper presented at the American Educational Research Association Annual Conference, San Diego, CA.
- Marion, S., & Fincher, M. (September 2008). *Validating a State's Alternate Assessment System*. 2008 Reidy Interactive Lecture Series, Portsmouth, NH.
- Grimm, C., Fincher, M., Link, V., Noland, M., & Teppoeva, T. (March 2008). *Portfolio Pattern Methodology: An Application for Alternate Assessment Standard Setting*. Paper presented at the American Educational Research Association Annual Conference, New York, NY.
- Chartrand, A., Cook, M., Davis, C., & Fincher, M. (June 2008). *1%, 2%: Skim or Whole Assessments*. 38th Annual National Conference on Student Assessment, Orlando, FL.
- Marion, S., Fincher, M., Stuck, J., Sims, R., Briggs, D., & Kearns, J. (June 2008). *Putting the Pieces Together: Validity Evaluations of States' Alternate Assessment Systems*. 38th Annual National Conference on Student Assessment, Orlando, FL.
- McDivitt, P. J., Felix, A., Fincher, M., Davis, C., & Quenemoen, R. (June 2008). *Redesign of Alternate Assessments for Students with Significant Cognitive Disabilities: National and State Perspectives*. 38th Annual National Conference on Student Assessment, Orlando, FL.
- Moen, R., Abedi, J., O'Brien, D., & Fincher, M. (June 2008). *Making Reading Tests Accessible by Affecting Student Test-taking Motivation*. 38th Annual National Conference on Student Assessment, Orlando, FL.
- Quenemoen, R., Fincher, M., Ban, P., & Draut, K. (June 2008). *Students Who are 'Difficult' to Assess: What Can We Do? How Will that Help?*. 38th Annual National Conference on Student Assessment, Orlando, FL.
- Fincher, M. (January 2008). *Understanding Who the Students Are: Using Data to Identify Eligible Students – State Perspective*. US Department of Education Kick-Off for GSEG Projects: Alternate Assessments Based on Modified Academic Achievement Standards, Washington, DC.
- Fincher, M. & Flowers, C. (October 2008). *Georgia's Alternate Assessment*. Paper presented at the Eighth Annual Maryland Assessment Conference, College Park, MD.

PROFESSIONAL SOCIETIES AND ORGANIZATIONS

National Council on Measurement in Education

ABSTRACT

INVESTIGATING THE EFFECTS OF A READ-ALoud ALTERATION ON THE THIRD-GRADE READING CRITERION-REFERENCED COMPETENCY TEST (CRCT) FOR STUDENTS WITH DISABILITIES

by
Melissa Fincher

The purpose of this study was to examine the effects of a controversial test administration alteration, the read-aloud alteration, in which text (passages and questions) is read aloud to the student on a reading comprehension test. For students whose disabilities impair their skill in decoding text and reading fluently, accessing text to demonstrate their comprehension can be significantly impeded. Using a quasi-experimental design, this study examined whether the comprehension scores for students with disabilities with certain characteristics improved with the read-aloud alteration. Participants were fourth-grade Georgia public school students (N=664) enrolled during the 2005-2006 school year, with and without disabilities, who were administered the third-grade Reading Criterion-Referenced Competency Test under either the read-aloud or standard administration condition. A 20-question survey was completed for each special education student who participated by the educator most familiar with the student's educational program. Several moderator variables, such as reading achievement as measured by an external criterion (the reading comprehension subtest of the Iowa Test of Basic Skills), the degree of the student's disability, as rated by the teacher, and individualized educational program features such as the presence of a decoding objective and time spent in the general education classroom, were investigated. These moderator variables were hypothesized to help better identify students with disabilities who might need and benefit from the read-aloud alteration. Students given

the read-aloud alteration achieved higher raw score gains on the posttest than those assessed under the standard condition regardless of their disability status (students with or without disabilities). No interactions were identified between the moderator variables studied and test condition, with the exception of testing condition (standard / read loud) and reading skill (below average, average, or above average). Regardless of disability status, students who were provided the read-aloud alteration and were classified as having below average reading skills on the norm-reference ITBS had higher gain scores than their peers.

INVESTIGATING THE EFFECTS OF A READ-ALoud ALTERATION ON THE
THIRD-GRADE READING CRITERION-REFERENCED COMPETENCY TEST
(CRCT) FOR STUDENTS WITH DISABILITIES

by
Melissa L. Fincher

A Dissertation

Presented in Partial Fulfillment of Requirements for the
Degree of
Doctor of Philosophy
in
Educational Psychology
in
the Department of Educational Psychology and Special Education
in
the College of Education
Georgia State University

Atlanta, GA
2013

Copyright by
Melissa Fincher
2013

ACKNOWLEDGEMENTS

Like most students, I have many people who supported me throughout my matriculation. My journey ended up being a rather long one during which quite a bit of life happened. I am grateful to my colleagues at the Georgia Department of Education, both present and former, who provided encouragement and rattled my cage when necessary. These include, but are not limited to, Sharron Hunt, Chris Domaleski, Jeff Barker, Dee Davis, Malina Monaco, Barbara Lunsford, and Tony Eitel.

Without a doubt I am indebted to my committee for their patience and belief in me. Without their encouragement, I would not have completed my studies. I particularly thank my chair, Dr. Commander. She was both gentle and kind while reminding me that the clock was ticking. Both Dr. Flowers and Dr. Roach prodded me to finish, with Dr. Flowers giving me the much needed permission to write trash, but write. And she was correct. Dr. Greenberg's confidence in me has been much appreciated.

Two of my fellow doctoral students – Ryan Hall and Kim Harrington Pete – provided invaluable support and motivation. Without their friendship I know I would not have finished. I owe them both so much.

I dedicate my work and my degree to my parents, Cameron and Mary Fincher. Both passed away before I was able to complete my degree but both were determined that I should finish. My father devoted his life's work to education in the state of Georgia, and my mother devoted her life to our family. I would not be the person I am today without their guidance and the influence of my siblings, Marcel, Matt, and Mandy. I appreciate their faith in me and love for me.

TABLE OF CONTENTS

	Page
List of Tables	v
List of Figures.....	vii
Abbreviations	viii
Chapter	
1 TEST ADMINISTRATION ACCOMMODATIONS AND MODIFICATIONS: A REVIEW OF THE LITERATURE.....	1
Need for Accommodations	1
Who are Students with Disabilities.....	6
Test Administration Accommodations	12
Methodologies Utilized to Investigate Accommodations.....	18
Differentiating Accommodations from Modifications	23
Differential Boost.....	26
Reliability.....	28
Factor Structure.....	29
Differential Item Functioning	31
Summary and Conclusions	33
References.....	38
2 INVESTIGATING THE EFFECTS OF A READ-ALoud ALTERATION ON THE THIRD-GRADE READING CRITERION-REFERENCED COMPETENCY TEST (CRCT) FOR STUDENTS WITH DISABILITIES.....	49
Students with Disabilities	53
Test Administration Accommodations	56
Read-Aloud Alteration.....	59

Methods.....	66
Participants.....	66
Procedure	70
Instruments.....	72
Data Analysis.....	75
Results.....	77
Disability Status and Level of Reading Skill.....	78
Degree of Disability.....	82
IEP Objective for Decoding.....	84
Time Spent in General Education Classroom.....	85
Grade Level of Instruction.....	87
Discussion.....	89
References.....	95
Appendixes	102

LIST OF TABLES

Table		Page
1	Percent of Students Reported in Each IDEA Disability Category	8
2	Summary of Reviewed Empirical Studies Investigating the Read-aloud Alteration	33
3	Percent of Students with Disabilities Reported in Each IDEA Category in 2008-2009	55
4	Demographic Characteristics of Fourth-grade Students by Disability Status and Test Condition.....	68
5	Number of Students with Disability in Each IDEA Category.....	69
6	Demographic Characteristics of Fourth-grade Students by Disability Status and Test Condition for ITBS-matched Sample	70
7	Means, Standard Deviations, Sample Size of Raw Score Gains by Disability Status and Level of Reading Skill	79
8	Three-way ANOVA Results for Test Condition by Disability Status and Level of Reading Skills.....	80
9	Effect Sizes for Reading Skill by Disability Status.....	82
10	Means, Standard Deviations, Sample Size of Raw Score Gains by Test Condition and Disability Status.....	83
11	Two-way ANOVA Results for Test Condition by Degree of Disability	84
12	Means, Standard Deviations, Sample Size of Raw Score Gains by Test Condition and Inclusion of Decoding Objective in IEP.....	85
13	Two-way ANOVA Results for Test Condition by Degree of Disability	85
14	Means, Standard Deviations, Sample Size of Raw Score Gains by Test Condition and Percent of Time Spent in General Education Classroom.....	86
15	Two-way ANOVA Results for Test Condition by Time Spent in General Education Classroom	87

16	Means, Standard Deviations, Sample Size of Raw Score Gains by Test Condition and Grade Level of Reading Instruction88
17	Two-way ANOVA Results for Test Condition by Grade Level of Reading Instruction88

LIST OF FIGURES

Figure		Page
1	The Interaction of Testing Condition by Level of Reading Skill for Students with Disabilities	81

ABBREVIATIONS

AA-AAS	Alternate Assessment – Alternate Achievement Standards
AA-MAS	Alternate Assessment – Modified Achievement Standards
ADA	Americans with Disabilities Act
AERA	American Educational Research Association
ANOVA	Analysis of Variance
APA	American Psychological Association
AYP	Adequate Yearly Progress
CRCT	Criterion-Referenced Competency Tests
DIF	Differential Item Functioning
IDEA	Individuals with Disabilities Education Act
IEP	Individualized Educational Program
ITBS	Iowa Test of Basic Skills
NARAP	National Accessible Reading Assessment Projects
NCE	Normal Curve Equivalency
NCLB	No Child Left Behind
NCME	National Council of Measurement in Education
NRP	National Reading Panel
SWD	Students with Disabilities
SWOD	Students without Disabilities
US ED	United States Department of Education

CHAPTER 1

TEST ADMINISTRATION ACCOMMODATIONS AND MODIFICATIONS: A REVIEW OF THE LITERATURE

Participation of *all* students in assessment and accountability systems is an important issue. The public relies on the information provided by assessment systems to gauge how well schools are doing in educating students. From a policy perspective, mandating the inclusion of all students, including those with disabilities, in assessment and accountability systems is meant to ensure the educational benefits afforded to most students are afforded to all students. Historically students with disabilities have not been given the same access to academic learning experiences and environments as their non-disabled peers. Until recent federal legislation and regulations, students with disabilities were exempted or excluded from the standardized academic assessments used in accountability models. The purpose of this paper is to review the growing body of literature surrounding test administration accommodations as a method of access for students with disabilities. Accommodations involve alterations to how a test is administered, presented, or how a student provides a response. Importantly, accommodations are designed to facilitate accurate measurement of students with disabilities' achievement. Although test administration accommodations may also be provided to eligible students with limited English proficiency, this literature review focuses on students with disabilities.

Need for Accommodations

Ensuring students with disabilities have the opportunity to learn and demonstrate their achievement is about equity. Historically, educational opportunity has been denied

to this group. Several federal laws, accompanying regulations, and court cases have guided the development of inclusive educational policy in regards to students with disabilities. Specifically, the Fourteenth Amendment of the United States Constitution, Section 504 of the Rehabilitation Act of 1973, the Education Act for All Handicapped Children Act of 1975, which later became the Individuals with Disabilities Education Act (IDEA), the Americans with Disabilities Act (ADA) of 1990 and No Child Left Behind (NCLB; 2002) have contributed to shaping educational policy in this area. In the context of inclusion, these laws are, in part, meant to ensure all students have equal educational opportunity (Giberson Schulte, Elliott, & Kratochwill, 2000; Phillips, 1994; Pitoniak & Royer, 2001).

The Rehabilitation Act of 1973 (Section 504) and the Americans with Disabilities Act of 1990 began to codify “societal orientation toward equity for individuals with disabilities” according to Gesinger (2007, p. ix). While the Education Act for All Handicapped Children Act of 1975 established the right for students with disabilities to receive a free and appropriate education, what constituted an *appropriate* education was left to interpretation. It was common practice to exempt or exclude students with disabilities from participation not only in grade-level instructional experiences, but also from academic assessments and accountability initiatives (Thurlow, 2007). The reasons for excluding students, particularly from assessments and accountability measures, ranged from the perception that the tests were not relevant for this population to protecting students from the frustration that would surely be prevalent during test administration, to concern and fear the anticipated poorer performance of students with disabilities would lower school scores (Elliott, 2007; Pitoniak & Royer, 2001).

According to Phillips (1994), IDEA was drafted as a remedy for the prior failure of public schools to appropriately serve and educate students with disabilities. She asserts, “IDEA was intended to provide educational services to disabled students who had been ignored, mistreated, or inappropriately institutionalized by the educational system” (p. 105). It was not until the reauthorization of IDEA in 1997 that the expectation was made explicit that students with disabilities should receive access to the same general academic curriculum as other students and participate in the same academic assessments. This edition of the law required that 1) all students with disabilities participate in state and district assessment and 2) their achievement be publicly reported in both aggregated form (with all students) and disaggregated form (as a subgroup). Even with that explicit expectation, however, the participation of students with disabilities in mandated assessment programs remained low (Fuchs, Fuchs, & Capizzi, 2005). No Child Left Behind (NCLB; 2002) further codified this expectation by requiring schools, districts, and states to release annual accountability determinations, called Adequate Yearly Progress (AYP), based on *all* students and specific subgroups, including students with disabilities. Only at this time did participation of students with disabilities increase. In short, schools, districts, and states are held accountable, under NCLB, for student participation in academic assessments as well as student performance (i.e., achievement). Schools must include a minimum of ninety-five percent of their students, at both the aggregated (all students) and disaggregated (subgroup) levels, in mandated academic assessments or they fail to make AYP on this criterion alone. Participation in assessments, as a result, has become high-stakes.

These mandates were designed to ensure access to the same academic content standards and learning opportunities experienced by their non-disabled peers (Phillips, 1994; Pullin, 2007; Roeber, 2002). According to Pullin (2007), the move toward inclusive policy was predicated on the belief that the students with disabilities would benefit from increased access to instruction based on the same grade-level standards taught to their peers and from accountability for the educational system that serves them. This is significant, because the historical exclusion of students with disabilities provided a false picture of how well a school was educating the students it served; in reality, an important portion of the student population was never given the opportunity to demonstrate their learning on assessments, distorting the means and distributions of scores (Elliott, 2007).

Both NCLB and IDEA provide options for including students with disabilities in assessment and accountability systems. The majority of students with disabilities are expected to participate in the general assessment program, that is, the same testing program general education students take, with or without test administration accommodations. Those students with the most significant cognitive disabilities, estimated to be approximately one percent of the general population or ten percent of the students with disabilities subgroup, may participate in an alternate assessment based on alternate achievement standards (AA-AAS). In 2007, the US Department of Education (US ED) introduced a fourth alternative for participation – an alternate assessment based on modified achievement standards (AA-MAS). This particular assessment option is also limited to students with disabilities and is capped by federal regulation at two percent of the general population (and consequently is often referred to as the 2% assessment).

In an effort to meet these requirements, states have struggled to provide access to grade-level curriculum and instruction, but access to standardized measures of student achievement has been particularly problematic. The initial efforts towards inclusion in assessment programs focused on participation. Test administration accommodations, changes in how tests are presented or administered or how students respond, were originally viewed as an avenue for achieving participation (Thurlow, 2007). As reauthorized in 1997, IDEA stipulated that students were to participate in assessment programs with appropriate accommodations as necessary. It failed to define, however, what constituted an appropriate accommodation, and as a result, how each state defined accommodations and which ones were considered allowable varied greatly (Chui & Pearson, 1999; Elliott, 2007; Fuchs & Fuchs, 1999; Hollenbeck, 2002).

Initially, little thought was given to what the tests were designed to measure and the impact of the accommodation(s) on the meaning or interpretation of the resulting test scores (Thurlow, 2007). Over time, policy and practice gradually shifted from an ‘anything goes’ approach to achieve participation towards ensuring the accommodations used by students with disabilities resulted in meaningful information about their achievement. Indeed, within the regulations for alternate assessments based on modified achievement standards, US ED (2007) clarified that students who participated in assessments through non-standard accommodations (also known as modifications), alterations in test administrations that interfere in some manner with the knowledge and skills (i.e., constructs) the test is designed to measure, could not be considered participants in AYP calculations. According to Zenisky and Sireci (2007), NCLB has encouraged careful consideration in both the policy and psychometric arenas about the

appropriateness of various test accommodations and their impact on test score interpretations.

Who are Students with Disabilities?

Before beginning the discussion of test administration accommodations, it is important to understand who students with disabilities are and why their access to standardized measures is an important issue within educational policy. According to Rooney (2011), in 2008-2009, approximately 6.5 million students, ages 3 – 21, received special education services. This represents 13 percent of the student population. Approximately 95% of these students were enrolled in public schools, with 57% spending the majority of their time in the general education classroom. (Note that IDEA stipulates that students with disabilities may be served until their 22nd birthday, much longer than their non-disabled peers.)

Determining just who is considered to have a disability is complicated by the fact that multiple federal laws offer slightly different definitions (Phillips, 1994, 2002; Pullin, 2007). In addition to the federal laws that require inclusion of students with disabilities in assessment and accountability systems, NCLB and IDEA, there are statutes that provide protection under civil rights. Specifically, both Section 504 of the Rehabilitation Act of 1973 and the ADA of 1990 prohibit discrimination on the basis of disability status (Pullin, 2007). Section 504 regulations require equal access and participation for any program or activity receiving federal funds. This regulation prohibits discrimination against otherwise qualified candidates because of their impairment(s) (Pitoniak & Royer, 2001). Title II of ADA prohibits exclusion, on the basis of an individual's disability, from participation, benefit, aid, or services for qualified individuals with disabilities.

ADA extends this safeguard to the private sector (Pitoniak & Royer, 2001). Section 504 (1973) and ADA (1990) define an individual with a disability as someone who has a disability that limits participation in major life activities, who is perceived as having a disability, or who has record of having had a disability. Because schooling can be considered one of life's major activities, both Section 504 and ADA can apply in the school setting (Pullin, 2007). Furthermore, the equal protection clause of the Fourteenth Amendment requires similarly situated students be treated equally (Phillips, 2002). The US Supreme Court, however, has clarified that the disabilities covered under these laws "must be substantial and do not include those in which there are measures for mitigating the impact of the disability, such as wearing eye glasses or contact lenses, or taking medication" (Pullin, 2007, p. 41).

Students with disabilities who receive special education services are a subset of the larger group covered under ADA and Section 504. For a student to be eligible for special education services under IDEA, the student 1) must have a disability and 2) there must be evidence that the disability impacts the student's learning such that there is a need for specialized services (Pullin, 2007). In other words, the mere existence of a disability is not sufficient in and of itself to warrant special education services. And although IDEA requires individualized educational plans outlining the specialized educational services to be provided a student with a disability, the law, as interpreted by the federal courts, does not guarantee any particular educational outcome (Geisigner, 1994; Phillips, 1994, 2002).

IDEA (2004) and its accompanying regulations define specific disability categories. These include: intellectual disabilities, hearing impairments (including

deafness), speech or language impairments, visual impairments (including blindness), emotional disturbance, orthopedic impairments, autism, traumatic brain injury, other health impairments, specific learning disabilities, and for children aged 3 – 9, developmental delay. Table 1 presents a breakdown of the percent of students classified into each disability category (Rooney, 2011).

Table 1

Percent of Students Reported in Each IDEA Disability Category

Disability Category	Percent of Students
Specific Learning Disabilities	42.9%
Speech Language Impairments	19.1%
Intellectual Disabilities	8.1%
Hearing Impaired	1.2%
Visual Impairment	0.4%
Emotional Disturbance	7.1%
Orthopedic Impairment	1.1%
Other Health Impairment	11.0%
Autism	5.0%
Traumatic Brain Injury	0.4%
Multiple Disabilities	2.1%

A disability can take many forms, including physical, sensory, or cognitive. Cognitive disabilities can include, but are not limited to, intellectual disabilities, dyslexia, dysgraphia, dyscalculia, and other learning disabilities. Further complicating this issue

is the fact that how a disability manifests itself within an individual is often unique. Additionally, the degree or severity of the same type of disability can vary greatly among individuals, and it is not uncommon for an individual to have multiple disabilities, which can interact. This makes the describing the characteristics of students with disabilities very difficult, as they comprise a highly heterogeneous group. Students with disabilities can have a wide assortment of cognitive and/or physical issues that present a multitude of academic challenges, ranging from mild to severe. As Pitoniak and Royer (2001) summarize, “there are great differences among individuals thus making any descriptions of the group as a single entity unadvisable” (p. 68).

Specific learning disabilities are the most prevalent category of disabilities reported in American schools (Pitoniak & Royer, 2001; Rooney, 2011; Thurlow, 2007). This category is broadly defined. It involves a disorder in the basic psychological processes of understanding and using language, whether spoken or written (IDEA, 2004). Learning disabilities can impact listening, thinking, speaking, reading, writing, spelling, and/or mathematical calculations, and can affect oral expression, listening comprehension, basic reading skills such as fluency and comprehension, as well as mathematical computation and problem solving skills. Students identified as having a specific learning disability comprise a diverse group, with varied degrees of academic deficits and strengths (Pitoniak and Royer, 2001). This is further complicated by the fact that what might be a strength for one student identified as having a learning disability may be a weakness for another (Fuchs, Fuchs, & Capizzi, 2005). In an effort to address this complexity, IDEA (2004) stipulates that a learning disability can only be identified when learners fail to benefit from effective, research-based instructional practices and

interventions designed to provide support. This is intended to help ensure that the deficit is truly related to the disability and is not a result of a lack of opportunity to learn based on the failure of the school to provide access to the curriculum or quality instruction.

Importantly, the nature of a student's disability can interact with the content and skills the student is expected to learn. Special education services, by their very premise, are designed to address the unique nature of each student's individual circumstance. Decisions about how to best serve students are made by a team comprised of the student's teacher(s), the student's parent(s), and the student him or herself, when appropriate. The Individualized Education Program (IEP) team is charged with developing a customized learning program, including providing additional supports and services as warranted, to address student needs. This team makes decisions about any necessary instructional and assessment accommodations in an effort to ensure students have access to content, are able to learn, and are situated to show what they have learned on academic measures, be they classroom tests or mandated large-scale achievement tests. IEP teams must understand the characteristics of the individual student's disability in order to make appropriate educational decisions that are in the best interest of the individual student. As Pullin (2007) describes, "the cornerstone of IDEA is the requirement that all students with disabilities receive an appropriate education, individually determined according to the student's IEP" (p. 41).

According to Hollenbeck (2005), the federal requirements assume IEP teams, specifically educators, have the knowledge and understanding necessary to make competent decisions, particularly when it comes to test administration accommodations. Gilbertson Schulte, Elliott, and Kratochwill (2000) found that the decision of which

accommodation(s) to provide was not influenced by the various policies and guidelines designed to inform such decisions, but rather that educators tend to make decisions based on their perceptions about accommodations. Particularly, educators consider whether an accommodation is believed to be helpful to the student and whether it is fair and feasible to implement. The study found that educators did not consider the severity of the disability. Lang et al. (2005) also found teachers tended to rely on those accommodations that could be used by many students rather than those tailored to individual need. Similarly, Hollenbeck (2005) reports that teachers have difficulty differentiating between students who would benefit from an accommodation from those who would not, as well as predicting which accommodation(s) would be helpful. Niebling and Elliott (2005) conclude that teacher judgment, as the sole criterion, may not be an appropriate approach for identifying appropriate accommodations, while Fuchs and Fuchs (1999) also urged caution based on their studies.

Thurlow (2007) reports that the number of empirical research studies investigating test administration accommodations and their impact on academic measures has increased since the federal requirements for inclusion and accountability were first introduced. Unfortunately, the lack of definitive guidance to help teachers and IEP teams make informed decisions about accommodations remains (Hollenbeck, 2002). Making appropriate decisions about accommodation is important to protect the integrity of the assessment and ensure accurate information about student achievement results. When test alterations are made, it is imperative to establish that the alteration neither overcorrects nor undercorrects in a manner that further distorts the student's performance thereby undermining the validity of the interpretations (Elliott, 2007). States continue to

struggle with developing comprehensive accommodation policies and guidance, based on empirical evidence and reasoned judgment, that ensure students with disabilities participate meaningfully in large-scale assessments (Cormier, Altman, Shyyan, & Thurlow, 2008).

Test Administration Accommodations

Test administration accommodations are used to provide students with disabilities an opportunity to demonstrate what they have learned and can do. Accommodations facilitate access to content and allow students with disabilities to show what they know and can do without the interference of their disability and its interaction with the content or test administration procedures (Cormier et al., 2008; Elliott, 2007; Hollenbeck, 2005; Phillips, 2002; Pitoniak & Royer, 2001; Sireci & Pitoniak, 2007; Thurlow, 2007).

According to the *Standards for Educational and Psychological Testing* (1999), jointly published by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), test administration accommodations are changes in the content, format, and/or administration procedure of a test in order to assist examinees who are unable to take the test under the standard conditions prescribed for the test. More succinctly, accommodations involve changes or alterations in the manner in which a test is administered, presented, or responded to by a student (Elliott, 2007; Thurlow, 2007; Niebling & Elliott, 2005). Pitoniak and Royer (2001) describe accommodations as customizing the materials and testing conditions based the examinee's needs. These changes are often categorized in terms of setting, timing, scheduling, presentation, and response.

Accommodations are intended to mitigate or lessen the impact of their disability but should not undermine the validity of the resulting test score interpretations (Hollenbeck, 2002; Sireci & Pitoniak, 2007). In other words, the purpose of an accommodation is to increase access as well as the accuracy of the measurement, and thereby the validity of the inference made from a test score. As described by the *Standards for Educational and Psychological Testing* (1999), the accommodations serve to “minimize the impact of test taker attributes that are not relevant to the construct that is the primary focus of the assessment” (p. 101).

Most large-scale assessments are standardized. Standardization involves ensuring uniform procedures, from test development and administration to scoring and reporting are followed, as a means of ensuring all parameters involved in assessment are the same (Geisinger, 1994; Green & Sireci, 1999; Phillips, 2002). In this manner, standardization provides surety that “any differences in student scores can be attributed to individual differences rather than to differences in testing procedure” (Sireci, Scarpati, & Li, 2005, p. 460). Often, it is the aspects of standardization that make the testing process difficult for students with disabilities. Accommodations serve, in part, to mitigate the impact or interaction of standardization with a disability, by allowing greater access so that the student may complete the test without confounding influences introduced by the test’s format, administration, or the manner in which a student provides a response (Hollenbeck, 2002). These types of confounding influences are unintended and introduce what Messick (1989) termed construct-irrelevant variance, a serious threat to validity. Construct-irrelevant variance involves the degree to which factors that are not germane to the construct (i.e., the knowledge and skills) the test is intended to measure are in fact

reflected in the test score. Such variance can be caused by the interactions of the student's disability with 1) the administration protocols (e.g., standardization) and/or 2) the actual content measured on the test.

Testing accommodations are designed to neutralize or remove construct-irrelevant variance caused by the disability (Chiu & Pearson, 1999; Geisinger, 1994; Sireci & Pitoniak, 2007; Sireci, Scarpati, & Li, 2005; Thurlow, 2007). Described by Chiu and Pearson (1999), accommodations can be likened to that of "a corrective lens" for potential score distortions caused by the student's disability and not their lack of knowledge and/or skill (p. 4). For example, allowing a student with a language processing disability extended time on a test of reading comprehension most likely does not change what the test is designed to measure, reading comprehension. The student is still engaged in reading but is allowed the compensation of additional, needed time to mitigate the impact of the disability. This assumes, of course, that the test is a power test focused on the level of achievement, rather than a speeded test focused on how quickly one can answer correctly within a specified time period.

It is important to note that accommodations are intended to maintain and facilitate access to the intended constructs of an assessment. In this manner, testing accommodations involve intentional changes or alterations to the manner in which a test is administered or how the student responds, but they do not involve changes in the content of the test, be it the stimuli (such as passages or scenarios) or the test items themselves (Elliott, 2007). Tindal (1998) defines an accommodation as a change that a) provides access so certain students may complete the tests and tasks without interference and more accurately demonstrate their achievement, but b) does not change the nature of

the content or skill being measured. Elliott, Braden, and White (2001) discuss accommodations in terms of target and access skills. Target skills are those specific concepts and skills the test is designed to measure (i.e., those skills or constructs targeted by the assessment, such as reading comprehension or mathematical problem solving). Access skills, on the other hand, are those that are used by the student to demonstrate the targeted skills (i.e., skills that provide student access, such as a quiet environment to concentrate or dictating the answer to a scribe). Access skills facilitate a student's participation in the assessment and help to eliminate measurement error due to poor or weak access skills. More succinctly, appropriate accommodations provide students support for deficits in access skills so that they can demonstrate their true achievement of the knowledge, concepts, and skills targeted by the assessment without the interference of the disability (Niebling & Elliot, 2005).

Accommodations differ from modifications, although the *Standards for Educational and Psychological Testing* (1999) use the two terms interchangeably and special educators tend to use the terms to signify the same meaning. Measurement specialists, however, define the terms differently and have worked over the last decade to clarify the differences. Tindal et al. (1997) describe a modification, also known as nonstandard accommodations, as a test alteration that changes the construct measured or one that works equally well for *all* students, therefore failing to provide differentiated support to students with specialized needs, such as those with disabilities. Modifications pose a threat to validity in that they introduce changes to the target skills, often resulting in different skills being measured. If the knowledge, content, or skills targeted by the assessment are somehow changed by the alteration employed to provide access to a

student, the resulting scores cannot be considered comparable. In other words, the score from a test with a modification is not comparable to a score from a standard or accommodated administration.

Consider for example, if a student with a disability is allowed to use a calculator on a test designed to measure computation. It would be difficult to draw the same inferences from the resulting test score as one would from a student who completed the assessment without the assistance of a calculator. While it may be possible for both students to achieve a score of 70% correct, it would be inaccurate to assume the scores provided an indication that the two students had comparable *computational* skills. In this example, the alteration in the test administration resulted in a fundamental change in the skill or construct targeted by the assessment, most likely resulting in an overestimation of the student's *computational* skill.

Elliott (2007) defines a modification as an alteration in test content, that is, in what the test measures, thus changing the validity of the inferences to be drawn from the results. Modifications challenge the degree to which we can be confident the test score is an accurate representation of a student's target skill level (Niebling & Elliott, 2005). Hollenbeck (2005) summarizes the differences between accommodations and modifications. The intended constructs of tests are upheld when accommodations are utilized. Modifications, however, alter fundamental elements of the test, which in turn change the content and the skills measured, potentially lowering the achievement expectation. Such a change calls into question the comparability of the test score and the inferences made from the score.

Hollenbeck (2005) proposes a continuum on which accommodations and modifications fall on opposite ends. He and his colleagues identified four attributes that must be present for a test alteration to be considered an accommodation. These include: a) unchanged constructs; b) individual need; c) differential effects; and d) sameness of inference. If all four are present, the alteration can be considered an accommodation. To ensure the appropriateness of a proposed alteration, it is important that it not change the concepts and skills the assessment is intended to measure. Its use should be predicated on a student need that has a relationship to the alteration and should provide a positive effect for students exhibiting that need but not for those who don't present such a need. Alterations that provide an advantage to all students fall on the modification end of the spectrum according to Tindal et al. (1997). Finally, in order for an alteration to be considered an accommodation, there must be comparability between scores of administrations that include the alteration and those that do not. The degree to which these four attributes are present can help determine where on the continuum between accommodation and modification a proposed alteration falls.

In her seminal article addressing the legal aspects of including students with disabilities in large-scale assessments, Phillips (1994) outlined a series of questions that should be asked to ascertain the appropriateness of a proposed alteration. She later revised these questions based on the growing understanding of the topic. In considering whether a test administration change is an accommodation or modification, Phillip (2002) suggests consideration of the following five questions:

1. Will the test score obtained under altered testing conditions have a different interpretation than scores obtained under standard test administration conditions?
Are the scores comparable?
2. Is the alteration in the test format or administration conditions part of the skill or knowledge being tested?
3. Would allowing the alteration for all students help nondisabled students achieve higher scores and change the interpretations of their test scores?
4. Can valid and reliable procedures and appeals be established for determining which students will be allowed which alterations?
5. Do students with disabilities included in regular education classrooms have any responsibility for adapting to standard testing conditions when feasible? (p. 125).

As she summarizes, “alternations in testing conditions fall on a continuum from little to no relationship to the skill being measured (an accommodation) to being significantly intertwined with the skill being assessed (a modification)” (p. 125).

Methodologies Utilized to Investigate Accommodations

In essence, the primary purposes of test administration accommodations are to promote access, equity, and validity for students with disabilities. Validity is at the heart of measurement and considers the degree to which information and evidence support the interpretations or inferences that are made from a test score. Importantly, validity is not a property of a test, but rather is concerned with the accuracy, fairness, and utility of the inferences, actions, or decisions that are made on the basis of test scores (Messick, 1989; Sireci and Green, 1999). *The Standards for Educational and Psychological Measurement* (1999) define validity as “the degree to which accumulated evidence and

theory support specific interpretations of test scores entailed by proposed uses of a test” (p. 184). In the context of accommodations, Barton (2007) suggests that validation must encompass multiple factors such as the testing environment, the administration protocols and procedures, the content standards measured and their relationship to the tested construct(s), the degree to which the test items singularly and as a group are representative of the intended construct(s), and student characteristics including, but not limited to, how the disabilities manifest themselves within students and interact with the construct(s).

Test developers and users have a responsibility to establish the utility and appropriateness of accommodations. Standard 10.1 in the *Standards for Educational and Psychological Testing* (1999) stipulates

In testing individuals with disabilities, test developers, test administrators, and test users should take steps to ensure that the test score inferences accurately reflect the intended construct rather than any disability and their associated characteristics extraneous to the intent of the measurement (p. 106).

This requirement is not easily achieved. Investigating test administration alterations to determine where they fall on the continuum, described previously as ranging from accommodation to modification, is a complicated endeavor. Researchers are faced with the challenge of quantifying the effects of alterations in an effort to determine whether proposed accommodations fulfill their purpose of increasing accessibility without changing what the test measures.

To establish the feasibility of a test alteration as an accommodation, both qualitative and quantitative methods must be employed. A fundamental step critical to

this process is ensuring the construct(s) measured by the test are clearly defined (Downing & Haladyna, 2004; Pullin, 2007; Phillips, 1994, 2002). Green and Sireci (1999) define a construct as a characteristic, skill, or ability that cannot be directly observed and that is believed to account for differences among individuals. Importantly, the definition of the construct(s) measured should be developed and articulated as part of the initial test development process. When developing accommodation policies for a test, Sireci and Pitoniak (2007) suggest convening a group of subject-matter experts to consider the effects of a proposed alteration on the knowledge, skills, and abilities the test is purported to measure. In this regard, potential accommodations should be conceptualized during the test development phase and consideration given to the interaction of the accommodation with the intended construct. Expert judgment should be made, when possible, to avoid alterations that interfere with or otherwise change the construct. While this is an important step, it is not sufficient in and of itself. Empirical evidence should also be gathered.

Historically, quantitative investigations have focused on establishing the utility of accommodations for students with disabilities. This involves establishing the benefit of the accommodation for the targeted students, those with disabilities, as opposed to providing an advantage to those without disabilities, often referred to as differential boost or the interaction hypothesis (Fuchs & Fuchs, 1999; Hollenbeck, 2000, 2002; Phillips, 1994; Sireci, Scarpati, & Li, 2005). Zuriff (2000) also referred to this as the maximum potential thesis (as cited in Sireci, Scarpati, & Li, 2005). As discussed earlier, Hollenbeck (2005) considers differential benefit one of the four criteria that must be satisfied in order to classify an alteration as an accommodation.

The interaction hypothesis stipulates that when test administration accommodations are matched correctly based on student need, test performance will improve compared to the scores obtained under standard conditions. Furthermore, and importantly, students without disabilities will not benefit from the accommodation (Sireci, Scarpati, & Li, 2005). This protocol typically requires a repeated measures 2 x 2 design and examines the interaction between student group membership (students with disability / students without disabilities) and test condition (accommodated / standard). As originally conceptualized, the interaction hypothesis was said to be met if the performance of students with disabilities improved under the accommodated condition when compared to their performance under standard conditions *and* the performance of students without disabilities remained consistent under both conditions (accommodated and standard). Sireci, Scarpati, & Li (2005) noted in their extensive literature review that often students without disabilities benefited from the accommodation as well, but not to the extent students with disabilities did. As a result, they suggested a revision, allowing gains for both groups of students but stipulated that gains for students with disabilities must be higher. According to Fuchs et al. (2000), differential boost requires students with disabilities must receive a significant test score increase compared to their nondisabled peers under the accommodated condition.

In general, three basic guidelines have guided the empirical investigations surrounding the effects of accommodations (Latusis, 2007; Randall & Engelhard, 2009). The first guideline, ensuring the accommodation provides a benefit to students in need, has already been introduced. The second and third involve investigating the impact of an alteration on the targeted construct and investigating the comparability of scores obtained

under accommodated conditions with those obtained under standard conditions. These two concepts, unaltered constructs and score comparability, are not mutually exclusive. A test score from an administration that used an alteration resulting in a changed construct is not comparable to a score from a standard administration; two different constructs have been measured. Similarly, if a student with a disability has not been provided an accommodation that is needed and accurate measurement is impeded by the disability, introducing construct-irrelevant variance, the resulting score is not comparable to one attained by a student without a disability. Empirical evidence of comparability can be investigated through experimental studies involving both students with and without disabilities and randomized treatment (alternation condition) assignment to students. And while empirical studies of this nature are considered the gold standard (Latusis, 2007), quasi-experimental, and non-experimental designs can also be informative (Sireci, Scarpati, & Li, 2005). Studies classified by the authors as quasi-experimental involve those cases in which the test administration conditions were manipulated but examinees were not randomly assigned. Non-experimental designs involved ex-post facto comparison of scores for students who received an accommodated test with those who received a standard administration.

Willingham et al. (1988) suggests several types of evidence test developers and psychometricians should collect to examine the comparability of scores and the impact of test alterations on the measured constructs of a test. Although this seminal work addresses testing accommodations for students with disabilities in the context of admissions testing, it has been applied to other types of assessment and has helped to establish the research protocols that are currently utilized today. Types of evidence

suggested by Willingham and his colleagues include examining how well the reliability, factor structure, and the functioning of items hold across test administration conditions and membership groups.

The need for test developers, which includes state departments of education, to establish the utility and appropriateness of allowable accommodations on mandated large-scale assessments has been pressed by the federal policies and legal implications discussed earlier. For example, to ensure the accuracy of measurement, the US ED (2007) requires states to develop clear accommodation policies and guidelines and document the technical quality of accommodated test scores. This evidence must be submitted for peer review. States must establish comparability between scores emanating from accommodated and standard administrations (Barton, 2007).

Differentiating Accommodations from Modifications

The following sections of this paper review empirical research studies published since 2000, the year in which states were mandated to include students with disabilities (IDEA, 1997). Because this literature review seeks to illuminate how test developers and psychometricians discern accommodations from modifications, the studies selected for review primarily focus on an alteration that is considered controversial, for which there is conflicting evidence and considerable debate within the measurement community. The read-aloud alteration permits students to have some or a test's entire text read aloud to facilitate their access. This alteration is typically considered for use by students who have disabilities that impact their language processing or reading skill. Use of the read-aloud or oral administration is considered controversial when used on tests that measure reading comprehension because it is believed to encroach on the construct measured.

Some advocates, however, argue that there are students whose disabilities prohibit their ability to access text and necessitate the need for the read-aloud or oral administration. They argue that unless decoding and fluency skills are explicitly contained within the test's constructs, often developed from state content standards, oral administration may be appropriate for *some* students (Laitusis, 2008, 2007).

Before reviewing the empirical studies found, it is important first to clearly outline what the literature says about the components of reading achievement. In 2000, after a comprehensive review of literature, the National Reading Panel (NRP) identified the key skills considered essential to reading achievement. These included phonemic awareness, phonics, fluency, vocabulary, and text comprehension. Phonemic awareness involves the understanding that the sounds of spoken language work to make words; that letters represent these sounds in written language. Phonics involves understanding the relationship between the letters of written language and the individual sounds of spoken language. Bridging word recognition and comprehension, fluency involves the ability to read text accurately and quickly. According to the NRP, fluent readers are able to both recognize and comprehend the meaning of a word at the same time. Likewise, vocabulary serves as a fundamental building block as readers must understand the meaning of individual words in order to comprehend. The NRP adopted Durkin's view that comprehension is the reason for reading. As such, reading is both purposeful and active (Durkin, 1993). The NRP defines reading comprehension as the construction of meaning from written text (National Institute of Child Health and Human Development, 2000).

Ultimately, reading is a cognitive process that integrates complex skills such as those identified by the NRP. In reporting the work of the National Accessible Reading Assessment Projects (NARAP), Thurlow (2010) suggests it is important to separate the individual components of reading into separate measures. She maintains that reading proficiency must include both comprehension and foundations skills such as decoding and fluency. No single component, such as phonemic knowledge, fluency, or comprehension, accounts for overall reading proficiency. She asserts, “students rely on their component proficiencies and the use of compensations to enable them to achieve overall proficiency in understanding a given text” (p. 124). By separating the components of decoding and fluency from comprehension, allowance of the read-aloud alteration may be permissible if the construct of interest for a test measuring comprehension is how well the student understands text. She cites college entrance exams such as the ACT and SAT as examples of testing programs that have defined the construct in this manner and allow the read-aloud as an accommodation for students with disabilities. According to Fielding and Pearson (1994), comprehension involves more than the literal recall of the author’s words; it also includes inferential and evaluative thinking.

Others argue that allowing text to be read aloud to students significantly changes the construct intended to be measured from reading comprehension to listening comprehension. The similarities and differences between reading and listening comprehension have been discussed for decades (Devine, 1968; Durrell, 1969; Tuman, 1980; Guthrie & Tyler, 1976). The fact that scores from reading and listening comprehension scores do not correlate that highly suggests that they may indeed be

tapping into different constructs (Devine, 1968). Thurlow (personal communication, November 22, 2011) agrees that the two skills are different. Listening comprehension has been defined as “receiving, attending to, interpreting, and responding to verbal messages” without text (US Department of Labor, 1991, p. 14). Thurlow (personal communication, November 22, 2011) points out that assessments of listening comprehension tend to involve a series of instructions, requesting the examinee follow those instructions, or having the examinee provide a missing word from a sentence presented orally using syntactic and semantic clues. She maintains listening comprehension assessments are designed to measure such things as (a) the degree to which the listener takes in raw speech and holds it in short term memory, (b) how the listener organizes what was heard into constituents, identifying their content and function, (c) when constituents are identified, how they are used to construct propositions, and how the propositions are grouped together to form a coherent message, and (d) how the identified and reconstructed propositional messages are held in long-term memory (Clark & Clarson, 1982, as cited by Thurlow, 2011).

In trying to ascertain the effect of the read-aloud alteration on test of reading comprehension, several studies were designed and conducted to investigate whether students with disabilities received benefit (differential boost) when provided this alteration, as well as to investigate the impact on the construct measured.

Differential Boost. Differential boost, as previously described, requires that students with disability receive a benefit from an alteration above any benefit received by their non-disabled peers. Five empirical research studies were identified that examined the effects of the read-aloud alteration on reading tests. The studies are described and the

findings are presented in order of the year they were conducted. Meloy et al. (2002) investigated the impact of the read-aloud alteration on four subtests, including the reading comprehension subtest, of the Iowa Test of Basic Skills (ITBS), a norm-referenced test. The study included both students with and without disabilities but did not use a repeated-measures design. Rather, students were randomly assigned to a condition (standard or read-aloud administration). Both students with and without disabilities benefited from the read-aloud alternation. Although students with disabilities benefited more, the interaction between condition and group membership was not significant.

McKevitt and Elliott (2003) investigated the effects of a read-aloud administration on a norm-referenced test with eighth grade students. Both students with and without disabilities took the test under two conditions (standard and read-aloud). The read-aloud consisted of an audiotape recording of the test's content. Results were not significant and no interaction or differential boost was detected for the read-aloud condition. Neither group of students performed better under the accommodated condition, although students without disabilities performed better than their peers with disabilities (which was true for standard condition as well).

Crawford and Tindal (2004) examined the effects of the read-aloud alteration on a reading comprehension test with fourth and fifth grade students. In this study, the read-aloud was administered through a video. A repeated measures design was employed and included both students with and without disabilities. Both groups benefited from the alteration; however, a significant interaction effect for students with disabilities was found. The authors issue a caution, however, about whether the use of this type of alteration on a reading test due to the constructs measured.

In an alteration of the standard read-aloud administration in which all text is read to students, Fletcher et al. (2006) examined the impact of reading proper nouns and the item question to students with and without dyslexia. A repeated measures design was also employed and yielded a significant interaction between administration type and student group membership. Students with poor decoding skills benefited the most from the read-aloud as operationalized in this study.

In a study involving both fourth and eighth graders, Laitusis (2010) found evidence of a differential boost in favor of students with reading-based learning disabilities. Although both groups of students, those with and without disabilities benefited from the read-aloud condition, students with disabilities benefited more.

Randall and Engelhard (2010a) found support for differential boost for third grade students, but not sixth grade students. In this repeated measures design, the performance of third grade students with disabilities improved with the read-aloud condition compared to their performance on the test during a standard administration. In grade 6, both student groups benefited.

Reliability. Reliability refers to the extent to which a test score is free of measurement error, providing an indication of the consistency or precision of a test. It is the degree to which an assessment yields consistent results across different items/tasks, times, settings, or raters. A measure that produces the same result when repeated is said to be reliable. Reliability also considers the consistency of items within a test; that is, when examining a test's internal structure, reliability provides an indication of the consistency of results across items on a test. It is important to examine the effects of a test alteration to ensure more error is not inadvertently introduced (Geisinger, 1994).

Given that the purpose of test accommodations is to reduce measurement error introduced by a student's disability, if an accommodation is effective, the reliability of the test should increase for students with disabilities when the accommodation is used. Randall and Engelhard (2010a) found the reliability of a high-stakes reading test increased for both groups of students (those with and without disabilities) when analyzing the effects of the read-aloud alteration for third grade students, although it increased more for students with disabilities. However, the opposite was found in grade 6. The reliability actually was slightly lower for both groups, but decreased more for students without disabilities under the accommodated condition than it did for students with disabilities.

Factor structure. Investigating the factor structure of accommodated test involves examining the intercorrelations among the test items that comprise the test. The pattern of correlations is believed to reveal the underlying factors that influence item performance. Items that are highly correlated are believed to be influenced by the same factor. In the arena of accommodations research, factor analysis is employed to help researchers evaluate whether the internal factor structure of a test remains similar across different examinee groups (students with disabilities / students without disabilities) and administration conditions (accommodated / standard). When the internal factor structure remains similar across examinee groups and conditions, measurement invariance is established (Randall and Engelhard, 2010a).

Randall and Engelhard (2010a) conducted a confirmatory factor analysis for a set of items comprising the Reading for Meaning domain on a high-stakes sixth-grade reading assessment. The authors found a unidimensional construct (i.e., one factor) when

comparing factor structure across group membership (students with or without disabilities) but not across administration condition (read-aloud or standard).

Cook et al. (2010) examined the factor structure of a fourth-grade English-language arts assessment. The study examined four groups of students. Students without disabilities who took the test under standard conditions, and students with learning disabilities who took the test under standard conditions, who took it with the accommodations specified in their IEP or 504 plans, and who took it with a read-aloud alteration. Although the results were not conclusive as one of the goodness-of-fit statistics yielded inconsistent results, the authors conclude that a similar factor structure held across all four groups.

Huynh and Barton (2006) investigated the effects of the read-aloud on a tenth-grade reading test. The study was comprised of three student groups. The first group consisted of students without disabilities who took the test under standard conditions, the second consisted of students with disabilities who were given the read-aloud alteration, and the third consisted of students with disabilities who were not given the alteration. The authors found that the internal structure of the assessment remained stable across student groups. Additionally, the internal structure remained stable across test condition when background variables were controlled.

Differential item functioning. Differential Item Functioning (DIF) involves investigating whether the probability of answering an item correctly differs across subgroups of examinees. Specifically, it identifies situations where the probability of answering an item correctly differs for two or more groups, who have been matched on ability. DIF occurs when group membership impacts the likelihood of responding

correctly to an item, which indicates that performance on the item may be determined by something other than the construct that is intended to be measured by the test. DIF procedures classify items ranging from negligible to intermediate to large. Qualitative consideration, by subject matter experts, is required to understand the reasons items may function differently across student groups or conditions.

Three recent studies analyzed item functioning on reading tests for which the read-aloud accommodations was provided. Bielinski et al. (2001) examined items on both a reading and mathematics test administered to elementary students in third and fourth grade. Both students with and without disabilities were included. Students without disabilities who took an unaccommodated test were matched in ability to students with disabilities who took the test with accommodations. Additionally two other groups were included, students with and without disabilities who did not receive the accommodation. DIF was identified for more items when the read-aloud accommodation was employed. While DIF was detected in both content areas when the read-aloud was provided, more items on the reading/language arts tests were flagged than on the mathematics test, suggesting the use of the accommodation impacts how some of the items function.

Bolt and Ysseldyke (2006) also used DIF to investigate the effects of the read-aloud accommodation for students with disabilities on both a reading and mathematics test. Three groups of students were analyzed: a random sample of students without disabilities (the reference group); students with disabilities who received the read-aloud alteration; and students with disabilities who received no accommodation. Far more DIF was detected in the read-aloud group on the reading/language arts assessment than on the

mathematics assessment, leading the authors to conclude that the read-aloud accommodation was associated with greater comparability concerns for reading tests than mathematics.

Randall & Engelhard (2010b) examined DIF for items measuring the domain of Reading for Meaning on a reading test in seventh grade as part of an investigation of the read-aloud accommodation. DIF analyses were conducted both at the group (with or without disabilities) and condition (standard or accommodated) levels. Minimal DIF was detected for one item, favoring students with disabilities, but no items were detected across administration type. The Reading for Meaning items functioned similarly when the read-aloud accommodation was provided as during the standard administration.

The results of the 12 studies identified, summarized in Table 2, often resulted in conflicting findings. For instance, three studies (Crawford & Tindal, 2004; Fletcher et al., 2006; Laitusis, 2010) found evidence that the read-aloud alteration provided a differential boost for students with disabilities, while two studies did not (McKevitt & Elliott, 2003; Meloy et al., 2002). Randall and Engelhard (2010a) found evidence of a differential boost in one grade, but not another. Likewise, evidence across the studies reviewed for internal structure and differential item functioning was mixed.

Unfortunately, mixed findings such as these are not an uncommon phenomenon (Thurlow, 2007). Generalizing the technical soundness of a test alteration across studies is difficult, if not impossible. Tests differ greatly in how they measure content and skills. Likewise, the populations included can differ. Often the findings are test and population specific but nonetheless inform the field on the utility and appropriateness of proposed alterations.

Table 2

Summary of Reviewed Empirical Studies Investigating Read-Aloud Alteration

<i>Study</i>	<i>Differential Boost</i>	<i>Reliability</i>	<i>Internal Structure</i>	<i>Differential Item Functioning</i>
Albedi et al. (2010)	N	Y	—	—
Bielinski et al. (2001)	—	—	—	N
Bolt & Ysseldyke (2006)	—	—	—	N
Cook et al. (2010)	—	—	Y	—
Crawford & Tindal (2004)	Y	—	—	—
Fletcher et al. (2006)	Y	—	—	—
Huynh & Barton (2006)	Y	—	Y	—
Laitusis (2010)	Y	—	—	—
Meloy et al. (2002)	N	—	—	—
McKevitt & Elliott (2003)	N	—	—	—
Randall & Engelhard (2010a)	P	P	—	—
Randall & Engelhard (2010b)	—	—	P	Y

Y= supporting evidence; P = partial supporting evidence; N = non-supporting evidence.

Summary and Conclusions

Although most schools have always worked to improve the services they provide, NCLB (2002) has compelled educators to search for ways to improve instruction and student learning with a renewed sense of urgency. This is particularly true for students with disabilities, who historically have not been given the same educational access or opportunities as their non-disabled peers and who have struggled to demonstrate what they have learned on standardized academic assessments. The purpose of this literature review was to examine the merits and technical quality of test administration accommodations provided to students with disabilities. Accommodations involve alterations to how a test is administered, are offered in an effort to help a student compensate for a disability that impedes their access to the content and skills measured,

and result in comparable scores and inferences. Accommodations are not provided to unfairly advantage certain students or to lessen learning or achievement expectations, but rather to help ensure measurement of student achievement is independent of disability. Modifications, on the other hand, alter what the test is designed to measure, have the potential to lower achievement expectations for students, and result in scores that are not comparable.

It is important that students with disabilities have an opportunity to learn and demonstrate their achievement just as their non-disabled peers are able to do. Unfortunately, as Fuchs and Fuchs (1999) assert, students with disabilities have been plagued with low expectations and limited opportunity. Access is an equity issue and as such, it is imperative for our educational system to consider the effects of the disability on the educational experiences of students (Geisinger, 1994) and to ensure students with disabilities are given the opportunity to learn the content and skills assessed. Test administration accommodations are designed to remove construct-irrelevant variance, a fundamental threat to validity, which can be introduced into the testing environment through a student's disability. As Sireci, Scarpati, and Li (2005) summarize, the primary psychometric agenda surrounding accommodations is maintaining construct representation while removing construct-irrelevant barriers to performance.

Development of accommodation protocols and policies is a critical step that begins in the test development process. Inappropriate use of an unproven or unfamiliar accommodation can introduce the very real possibility that observed score differences are the result of variation in the administration procedures and not differences in the underlying constructs (Geisinger, 1994). As Hollenbeck (2002) writes, "paradoxically,

although accommodations are utilized to reduce measurement error, they may interject additional sources of error into the assessment” (p. 401). Navigating the tensions between individualization, standardization, and accountability present a myriad of challenges for test developers, psychometricians, and policy makers (Pullin, 2007). Students with disabilities are entitled to specialized, individualized services to address the unique needs brought on by their disability. This entitlement often conflicts with the standardization required by assessment and accountability systems.

Several factors make the study of test administration accommodations challenging. As discussed previously, student disabilities are often unique, with the same disability manifesting itself differently within individuals. This is further complicated by the fact that for several disability categories the numbers are small, which limits the types of analyses that can be conducted (Pitoniak & Royer, 2001; Geisinger, 1994). Establishing a sufficient sample of students with similar needs is not easy and often prohibitive. Likewise, it can be impractical to conduct rigorous experimental studies with randomized assignment of conditions within the public education system. According to Thurlow (2007), developing a “research-based decision-making practice in which accommodation issues are identified by a state and then researched through randomized trial designs to produce a decision about accommodation policy may have to remain an ideal” (p. 19).

Generalizability across studies is challenging given the differences in measures (i.e., tests) studied, the populations sampled, and the implementation of the alterations. Barton (2007) also maintains that there is significant variability in the identification of students with disabilities and the assignment of disability categories. This is

compounded by the fact that many students have multiple disabilities and most students receive multiple accommodations, administered in a bundle. As Malouf (2005) summarizes, studies that group students with like disabilities or combine different accommodations are less informative than studies that focus on a specific accommodation and define the student group based on common characteristics of the disability the accommodation is intended to mitigate.

At a recent technical assistance meeting hosted by the US Department of Education, Albedi (2011) recommended six points to focus and improve research on accommodations. These include effectiveness, validity, differential impact, comparability, relevance, and feasibility. Several of these points have been discussed in this paper. Albedi, however, contends that research would be improved if studies also examined the how effective and relevant proposed alterations were in terms of increasing accessibility for targeted students as well as considered how feasible they were to implement. Thurlow (2011) suggests a need for improved selection of students for participation in research studies, ensuring studies focus on students who need and would potentially benefit from the accommodation. As both Fletcher et al. (2006) and Tindal et al. (1998) have pointed out, often the research conducted to date does not specifically target the students who need the accommodation in the content area studied. “To provide the most convincing empirical support for an accommodation, students with a specific need have to be compared to others without such a need who are otherwise comparable in achievement” (Tindal et al., 1998, p.442).

Advancing from individual research studies to an integrated, coherent program of validity research on test comparability when accommodations are employed would

represent substantial progress (e.g., Latusis, 2007; Thurlow, 2007). With the development of common assessments by consortia of states, the opportunity to advance research on accommodations and address many of the technical deficits noted is, perhaps for the first time, at hand. These multi-state consortia have the opportunity to carry out a robust research agenda on a common assessment, with a common construct(s). While the variability in how students are identified for special education services and assigned a particular disability category label may remain, sampling can more easily focus on identifying student needs within a content area and not overtax a single state, district, or school.

References

- Albedi, J. (2011, August). What is the current state of research on accommodations and what future research is necessary? Presentation at the Race to the Top Assessment Program Technical Assistance Public Meeting for Creating Valid, Reliable, and Fair Assessments for Students with Disabilities and English Learners , Washington, D.C.
- All Handicapped Children Act, 10, U.S.C. 1400 *et seq.* (1975).
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standard for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Americans with Disabilities Act of 1990, 42 U.S.C. § 12101 *et seq.*
- Barton, K. E. (2007). Validity and accommodations: The journey toward accessible assessments. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 81-94). Arlington, VA: Council for Exceptional Children.
- Bielinski, J., Thurlow, M., Ysseldyke, J. E., Friedebach, J., & Friedebach, M. (2001). *Read-aloud accommodation: Effects on multiple-choice reading and math items* (Tech. Rep. No. 31). Minneapolis, MN: University of Minnesota, National Center for Educational Outcomes.
- Bolt, S. E., & Thurlow, M. L. (2004). Five of the most frequently allowed testing accommodations in state policy: Synthesis of research. *Remedial and Special Education, 25*(3), 141-152.

- Bolt, S., & Ysseldyke, J. (2006). Comparing DIF across math and reading/language arts tests for students receiving a read aloud accommodation. *Applied Measurement in Education, 19*(4), 329-355.
- Braden, J. P., Huai, N., White, J. L., & Elliott, S. N. (2005). Effective professional development to support inclusive large-scale assessment practices for all children. *Assessment for Effective Interventions, 31*, 1, 63-72.
- Chiu, C. WT, & Pearson, P. D. (1999, June). *Synthesizing the effects of test accommodations for special education and limited English proficient students*. Paper presented at the National Conference on Large-Scale Assessment, Snowbird, UT.
- Clark, H.H., & Carlson, R. (1982). Hearers and speech acts. *Language, 58* (2), 332-373.
- Cormier, D. C., Atلمان, J. R., Shyyan, V., & Thurlow, M. L. (2010). A summary of the research on the effects of test accommodations: 2007-2008 (Technical Report 56). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Crawford, L. & Tindal, G. (2004). Effects of a read-aloud accommodation on a standardized reading test. *Exceptionality, 12*(2), 89-106.
- Devine, T. (1968). Reading and listening: New research finding. *Elementary English, 45*, 346-348.
- Downing, S. M. & Haladyna, T. M (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement in Education, 10*(1), 61-82.
- Durrell, D. D. (1969). Listening comprehension versus reading comprehension. *Journal*

of Reading, 12(6), 455-460.

Education for All Handicapped Children Act, 20 U.S.C. § 1400 et seq (1975).

Elbaum, B. (2007). Effects of an oral testing accommodation on the mathematics performance of secondary students with and without learning disabilities. *The Journal of Special Education*, 40(4), 218-229.

Elbaum, B., Arguelles, M., Campbell, Y., & Saleh, M. (2004). Effects of a student-reads-aloud accommodation on the performance of students with and without learning disabilities on a test of reading comprehension. *Exceptionality*, 12(2), 71-87.

Elliott, S. N. (2007). Selecting and using testing accommodations to facilitate meaningful participation of all students in state and district assessments. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 1-12). Arlington, VA: Council for Exceptional Children.

Elliott, S. N., Braden, J. P., & White, J. L. (2001). *Assessing one and all: Educational accountability for students with disabilities*. Arlington, VA: Council for Exceptional Children.

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(2), 136-150.

Fuchs, L. S. & Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research and Practice*, 16, 174-181.

Fuchs, L. S., & Fuchs, D. (1999). Fair and unfair testing accommodations. *School*

Administrator, 56, 24-29.

- Fuchs, L. S., Fuchs, D., & Capizzi, A. M. (2005). Identifying appropriate test accommodations for students with learning disabilities. *Focus on Exceptional Children*, 37(6), 1-8.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about testing accommodations. *Exceptional Children*, 67, 67-81.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., & Karns, K. (2000). Supplementing teacher judgments of mathematics test accommodating with objective data sources. *School Psychology Review*, 29, 65-85.
- Geisigner, K. F. (2007). Preface. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. ix-xii). Arlington, VA: Council for Exceptional Children.
- Geisigner, K. (1994). Psychometric issues in testing students with disabilities. *Applied Measurement in Education* 7(2), 121-141.
- Gibson, D., Haeberli, F. B., Glover, T. A., & Witter, E. A. (2005). Use of recommended and provided testing accommodations. *Assessment for Effective Interventions*, 31, 1, 19-36.
- Gilberston Schulte, A. A., Elliott, S. N., & Kratochwill, T. R. (2000). Educators' perceptions and documentation of testing accommodations for students with disabilities. *Special Services in the Schools*, 16, 35-56.
- Green, P. C., & Sireci, S. G. (1999). Legal and psychometric policy considerations in the testing of students with disabilities. *Journal of Special Education*

Leadership, 12(2), 21-29.

Guthrie, J., & Tyloer, S. J. (1976). Psycholinguistic processing in reading and listening among good and poor readers. *Journal of Reading Behavior*, 8, 415-425.

Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children*, 69, 211-225.

Hollenbeck, K. (2005). Validity issues and decisions about testing accommodations. *Assessment for Effective Interventions*, 31, 1, 7-18.

Hollenbeck, K. (2002). Determining when test alterations are valid accommodations or modifications for large-scale assessments. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, Technical Adequacy, and Implementation* (pp. 395-425). Mahwah, NJ: Lawrence Erlbaum Associates.

Hollenbeck, K., Tindal, G. & Almond, P. (1998). Teachers' knowledge of accommodations as a validity issue in high-stakes testing. *Journal of Special Education*, 32, 175-183.

Huynh, H., & Barton, K. (2006). Performance of students with disabilities under regular and oral administrations of a high-stakes reading examination. *Applied Measurement in Education*, 19(1), 21-39.

Huynh, H., Meyer, J. P., & Gallant, D. J. (2004). Comparability of student performance between regular and oral administrations for high-stakes mathematics tests. *Applied Measurement in Education*, 19(1), 21-39.

Individuals with Disabilities Education Act (IDEA), as amended in 2004, PL 108-446, 20 USC 1400 et seq.

Individuals with Disabilities Education Act of 1997, Pub. L. No. 105-17, 111 Stat. 37

(1997).

Kettler, R. J., Niebling, B. C., Mroch, A. A., Feldman, E. S., Newell, M. L., Elliott, S.N., Kratochwill, T. R., & Bolt, D. M. (2005). Effects of testing accommodations on math and reading scores: An experimental analysis of the performance of students with and without disabilities. *Assessment for Effective Interventions, 31*, 1, 37-48.

Ketterlin-Geller, L. R., Yovanoff, P., & Tindal, G. (2007). Developing a new paradigm for conducting research on accommodations in mathematics testing. *Exceptional Children, 73*(3), 331-347.

Lang, S. C., Kumke, P. J., Ray, C. E., Cowell, E. L., Elliott, S. N., Kratochwill, T. R., & Bolt, D. M. (2005). Consequences of using testing accommodations: Student, teacher, and parent perceptions of and reactions to testing accommodations. *Assessment for Effective Interventions, 31*, 1, 49-62.

Laitusis, C. C. (2010). Examining the Impact of Audio Presentation on Tests of Reading Comprehension. *Applied Measurement in Education, 23*, 153-167.

Laitusis, C. C. (2008). State reading assessments and inclusion of students with dyslexia. *Perspectives on Language and Literacy, 34*(4), 31-33.

Laitusis, C. C. (2007). Research designs and analysis for studying accommodations on assessments. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 67-80). Arlington, VA: Council for Exceptional Children.

Malouf, D. B. (2005). Accommodation research to guide practice: Comments on

what we know and what we need to know. *Assessment for Effective Interventions*, 31, 1, 79-83.

McKevitt, B. C., & Elliott, S. N. (2003). Effects and perceived consequences of using read-aloud and teacher-recommended accommodations on a reading achievement test. *School Psychology Review*, 32, 583-600.

Meloy, L. L., Deville, C., & Frisbie, D. A. (2002). The effects of a read-aloud accommodation on test scores of students with and without learning disability in reading. *Remedial and Special Education*, 23(4), 248-255.

Messick, S. (1989). Validity. In R. Linn (Ed.) *Educational measurement* (3rd ed., pp. 13-103). Washington, DC: American Council on Education.

National Institute of Child Health and Human Development. (2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: Author.

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

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Phillips, S. E. (2002). Legal issues affecting special populations in large-scale testing programs. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, Technical Adequacy, and Implementation* (pp. 395-425). Mahwah, NJ: Lawrence Erlbaum Associates.

Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education*, 7, 2, 93-120.

- Pitoniak, M. J., & Royer, J. M. (2001). Testing accommodations for examinees with disabilities: A review of psychometric, legal and social policy issues. *Review of Educational Research, 71*(1), 53-104.
- Pullin, D. (2007). Accommodations in testing: Law, policy, and practice. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 37-50). Arlington, VA: Council for Exceptional Children.
- Randall, J. & Engelhard, G. (2010a). Performance of students with and without disabilities under modified conditions. *The Journal of Special Education, 44*(2), 79-93.
- Randall, J., & Engelhard, G. (2010b). Using confirmatory factor analysis and the Rasch model to assess measurement invariance in a high stakes reading assessment. *Applied Measurement in Education, 23*, 286-306.
- Roeber, E. D. (2002). Appropriate inclusion of students with disabilities in state accountability systems. Denver, CO: Education Commission of the States.
Retrieved June 30, 2005, from World Wide Web:
<http://www.ecs.org/clearinghouse/40/11/4011.htm>.
- Rooney, P. (2011, August). Understanding the population. Presentation at the Race to the Top Assessment Program Technical Assistance Public Meeting for Creating Valid, Reliable, and Fair Assessments for Students with Disabilities and English Learners , Washington, D.C.
- Section 504 of the Rehabilitation Act of 1973, 29 U.S.C. § 701 *et seq.*
- Sireci, S. G., & Pitoniak, M. J. (2007). Assessment accommodations: What have we

- learned from research? In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 53-67). Arlington, VA: Council for Exceptional Children.
- Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research, 75*(4), 457-490
- Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (2001). Effects of testing accommodations on standardized mathematics test scores: An experimental analysis of the performances of students with and without disabilities. *School Psychology Review, 30*, 527-547.
- Shriner, J. G. (2000). Legal perspectives on school outcomes assessment for students with disabilities. *Journal of Special Education, 30*(4), 232-239.
- Thurlow, M. (2011, August). What is the current state of research on accommodations and what future research is necessary? Presentation at the Race to the Top Assessment Program Technical Assistance Public Meeting for Creating Valid, Reliable, and Fair Assessments for Students with Disabilities and English Learners , Washington, D.C.
- Thurlow, M. L. (2011). Steps towards creating fully accessible reading assessments. *Applied Measurement in Education, 23*, 121-131.
- Thurlow, M. L. (2007). State policies and accommodations: Issues and implications. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 13-22). Arlington, VA: Council for Exceptional Children.

- Tindal, G. (1998). *Models for understanding task comparability in accommodated testing*. Eugene, OR: Behavioral Research and Teaching.
- Tindal, G., & Fuchs, L. S. (1999). *A summary of research on test changes: An empirical basis for defining accommodations*. Lexington: University of Kentucky, Mid-South Regional Resource Center Interdisciplinary Human Development Institute.
- Tindal, G., Heath, B., Hollenbeck, K., Almond, P., & Harniss, M. (1998). Accommodating students with disabilities on large-scale tests: An experimental study. *Exceptional Children*, 64, 439-450.
- Tuman, M. C. (1980). A comparative review of reading and listening comprehension. *Journal of Reading*, 23(8), 698-704.
- US Department of Education, Office of Elementary and Secondary Education. (2007a). *Standards and assessments peer review guidance: Information and examples for meeting the requirements of no child left behind act of 2001*. Washington, DC: Author.
- US Department of Education, Office of Elementary and Secondary Education. (2007b). *Modified achievement standards: Non-regulatory guidance*. Washington, DC: Author.
- US Department of Labor. (1991). *What work requires of schools (A SCANS report for America 2000)*. Washington, DC: Author. Available at <http://wdr.doleta.gov/SCANS/whatwork/whatwork.pdf>.
- Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, H., Rock, D. A., Powers, D. E. (1988). *Testing handicapped people*. Needham, MA: Allyn & Bacon.
- Zenisky, A.L. & Sireci, S.G. (2007). *A Summary of the Research for the Effects of Test*

Accommodations: 2005 through 2006. Center for Educational Assessment,
University of Massachusetts-Amherst.

CHAPTER 2
INVESTIGATING THE EFFECTS OF A READ-ALoud ALTERATION ON THE
THIRD-GRADE READING CRITERION-REFERENCED COMPETENCY TEST
(CRCT) FOR STUDENTS WITH DISABILITIES

Participation of all students in assessment and accountability systems is an important issue. The public relies on the information provided by these systems to gauge how well schools are doing in educating students. From a policy perspective, mandating the inclusion of all students, including those with disabilities, in assessment and accountability systems is meant to ensure educational benefits are afforded to all students. Historically students with disabilities have not been given the same access to academic learning experiences and environments as their non-disabled peers, and as a result, educational opportunity has been denied this group (Phillips, 2002, 1994; Pullin, 2007; Thurlow, 2007). Until federal legislation and regulations mandated their inclusion and participation beginning in 2000, students with disabilities were frequently exempted or excluded from the standardized academic assessments used in accountability models.

Ensuring students with disabilities have the opportunity to learn and demonstrate their achievement is about equity (Pullin, 2007). Several federal laws and court cases have guided the development of inclusive educational policy for students with disabilities. Specifically, the Fourteenth Amendment of the United States Constitution, Section 504 of the Rehabilitation Act of 1973, the Education Act for All Handicapped Children Act of 1975, which later became the Individuals with Disabilities Education Act (IDEA), the Americans with Disabilities Act (ADA) of 1990 and No Child Left Behind (NCLB; 2002) contributed to shaping educational policy in this area (Phillips, 2002;

Pullin, 2007). In the context of inclusion these laws are, in part, meant to ensure all students have equal educational opportunity (Gilbertson Schulte, Elliott, & Kratochwill, 2000; Phillips, 1994; Pitoniak & Royer, 2001). According to Gesinger (2007), the Rehabilitation Act of 1973 (Section 504) and the ADA of 1990 began to codify “societal orientation toward equity for individuals with disabilities” (p. ix).

While the Education Act for All Handicapped Children Act of 1975 established the right for students with disabilities to receive a free and appropriate education, what constituted an *appropriate* education was left to interpretation. It was common practice to exempt or exclude students with disabilities from participation not only in grade-level instructional experiences, but also from academic assessments and accountability initiatives (Thurlow, 2007). The reasons for excluding students, particularly from assessments and accountability measures, ranged from the perception that the tests were not relevant for this population to protecting students from the frustration that would surely be prevalent during test administration, to concern and fear the anticipated poorer performance of students with disabilities would lower school scores (Elliott, 2007; Pitoniak & Royer, 2001).

According to Phillips (1994), IDEA was drafted as a remedy for the prior failure of public schools to appropriately serve and educate students with disabilities. She asserts, “IDEA was intended to provide educational services to disabled students who had been ignored, mistreated, or inappropriately institutionalized by the educational system” (p. 105). It was not until the reauthorization of IDEA in 1997 that the expectation was made explicit that students with disabilities should receive access to the same general academic curriculum as other students and participate in academic assessments. This

version of the law required that 1) all students with disabilities participate in state- and district-mandated assessments and 2) their achievement be publicly reported in both aggregated form (with all students) and disaggregated form (as a subgroup). However, even with this explicit expectation, the participation of students with disabilities in mandated assessment programs remained low (Fuchs, Fuchs, & Capizzi, 2005). NCLB (2002) further codified this expectation by requiring schools, districts, and states to release annual accountability determinations, called Adequate Yearly Progress (AYP), based on *all* students and specific subgroups, including students with disabilities. Only after the inclusion of this requirement in federal policy did participation of students with disabilities increase. In short, schools, districts, and states are held accountable, under NCLB, for student participation in academic assessments as well as student performance (i.e., achievement). Schools must include a minimum of 95% of their students, at both the aggregate (all students) and disaggregated (subgroup) levels, in mandated academic assessments or they fail to make AYP on the participation criterion alone.

As previously mentioned, these mandates were designed to ensure access to the same academic content standards and learning opportunities experienced by their peers without disabilities (Phillips, 1994; Pullin, 2007; Roeber, 2002). According to Pullin (2007), the move toward inclusive policy was predicated on the belief that students with disabilities will benefit from increased access to instruction based on the same grade-level standards taught to their peers and from accountability for the educational system that serves them. This is important because the past exclusion of students with disabilities provided a false picture of how well a school was educating the students it served. In reality, a major portion of the student population was never given the

opportunity to demonstrate their learning on assessments, distorting the means and distributions of test scores presented to the public as evidence of a school's educational record (Elliott, 2007).

In an effort to meet these requirements, states have struggled to provide access to grade-level curriculum and instruction, with access to standardized measures of student achievement being particularly problematic. The early efforts towards inclusion in assessment programs focused on participation. Test administration accommodations, changes in how tests are presented or administered or how students respond, were originally viewed as an avenue for achieving participation (Thurlow, 2007). As reauthorized in 1997, IDEA stipulated that students were to participate in assessment programs with appropriate accommodations as necessary. It failed to define, however, what constituted appropriate accommodations, and as a result, how each state defined accommodations varied greatly (Chui & Pearson, 1999; Elliott, 2007; Fuchs & Fuchs, 1999; Hollenbeck, 2002).

Initially, little thought was given to what the tests were designed to measure and the impact of the accommodation(s) on the meaning or interpretation of the resulting test scores (Thurlow, 2007). Over time, policy and practice gradually shifted from an 'anything goes' approach to achieve participation towards ensuring the accommodations used by students with disabilities resulted in meaningful information about their achievement. Indeed, within the regulations for alternate assessments based on modified achievement standards, US ED (2007) clarified that students who participated in assessments through non-standard accommodations (also known as modifications), alterations in test administrations that interfere in some manner with the knowledge and

skills the test is designed to measure, could not be considered participants in AYP calculations. According to Zenisky and Sireci (2007), NCLB has encouraged careful consideration in both the policy and psychometric arenas about the appropriateness of various test accommodations and their impact on test score interpretations.

Students with Disabilities

It is important to understand who students with disabilities are and why their access to standardized measures is an important issue within educational policy. According to Rooney (2011), in 2008-2009, approximately 6.5 million students ages 3 – 21, received special education services. (Note that IDEA stipulates that students with disabilities may be served until their 22nd birthday, much longer than their non-disabled peers.) This represents 13 percent of the student population. Approximately 95% of these students were enrolled in public schools, with 57% spending the majority of their time in the general education classroom.

Determining just who is considered to have a disability is complicated by the fact that multiple federal laws offer slightly different definitions (Phillips, 1994, 2002; Pullin, 2007). Section 504 (1973) and ADA (1990) define an individual with a disability as someone who has a disability that limits participation in major life activities, who is perceived as having a disability, or who has record of having had a disability. Because schooling can be considered one of life's major activities, both Section 504 and ADA can apply in the school setting (Pullin, 2007). Furthermore, the equal protection clause of the Fourteenth Amendment requires similarly situated students be treated equally (Phillips, 2002). The United States Supreme Court, however, has clarified that the disabilities covered under these laws "must be substantial and do not include those in which there are

measures for mitigating the impact of the disability, such as wearing eye glasses or contact lenses, or taking medication” (Pullin, 2007, p. 41).

Students with disabilities who receive special education services are a subset of the larger group covered under ADA and Section 504. For a student to be eligible for special education services under IDEA, 1) the student must have a disability and 2) there must be evidence that the disability impacts the student’s learning such that there is a need for specialized services (Pullin, 2007). In other words, the mere existence of a disability is not sufficient in and of itself to warrant special education services.

IDEA (2004) and its accompanying regulations define specific disability categories. These include intellectual disabilities, hearing impairments (including deafness), speech or language impairments, visual impairments (including blindness), emotional disturbance, orthopedic impairments, autism, traumatic brain injury, other health impairments, specific learning disabilities, and for children aged 3 – 9, developmental delay. Table 3 presents a breakdown of the percent of students classified into each disability category (Rooney, 2011).

A disability can take many forms, including physical, sensory, or cognitive. Cognitive disabilities can include, but are not limited to, intellectual disabilities, dyslexia, dysgraphia, dyscalculia, and other learning disabilities. Furthermore, how a disability manifests itself within an individual is often unique. The degree or severity of the same type of disability can vary greatly among individuals, and it is not uncommon for an individual to have multiple disabilities which can interact. These circumstances make describing the characteristics of students with disabilities very difficult, as they comprise a highly heterogeneous group. In reality, students with disabilities can have a wide

assortment of cognitive and/or physical issues that present a multitude of academic challenges, ranging from mild to severe. Pitoniak and Royer (2001) issue caution, stating “there are great differences among individuals thus making any descriptions of the group as a single entity inadvisable” (p. 68).

Table 3

Percent of Students with Disabilities Reported in Each IDEA Category in 2008-2009

Disability Category	Percent
Specific Learning Disabilities	42.9%
Speech Language Impairments	19.1%
Intellectual Disabilities	8.1%
Hearing Impaired	1.2%
Visual Impairment	0.4%
Emotional Disturbance	7.1%
Orthopedic Impairment	1.1%
Other Health Impairment	11.0%
Autism	5.0%
Traumatic Brain Injury	0.4%
Multiple Disabilities	2.1%

Importantly, the nature of a student’s disability can interact with the content and skills the student is expected to learn. Special education services, by their very premise, are designed to address the unique nature of each student’s circumstance. Decisions about how to best serve students are made by a team comprised of the student’s teacher(s), the student’s parent(s), and the student him or herself, when appropriate. The

Individualized Education Program (IEP) team is charged with developing a customized learning program, providing additional supports and services as warranted, to address student needs. This team makes decisions about any necessary instructional and assessment accommodations in an effort to ensure students have access to content, are able to learn, and are situated to show what they have learned on academic measures, be they classroom tests or mandated large-scale achievement tests. IEP teams must understand the characteristics of the individual student's disability in order to make appropriate educational decisions that are in the best interest of the individual student. As Pullin (2007) describes, "the cornerstone of IDEA is the requirement that all students with disabilities receive an appropriate education, individually determined according to the student's IEP" (p. 41).

Test Administration Accommodations

Test administration accommodations are used to provide students with disabilities an opportunity to demonstrate what they have learned and can do. Accommodations facilitate access to content and allow students with disabilities to show what they know and can do without the interference of their disability and its interaction with the content or test administration procedures (Cormier et al, 2008; Elliott, 2007; Hollenbeck, 2005; Phillips, 2002; Pitoniak & Royer, 2001; Sireci & Pitoniak, 2007; Thurlow, 2007). According to the *Standards for Educational and Psychological Testing* (1999), jointly published by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), test administration accommodations are changes in the format and/or administration procedure of a test in order to assist examinees who are unable to take the

test under the standard conditions prescribed for the test. More succinctly, accommodations involve changes or alterations in the manner in which a test is administered, presented, or responded to by a student (Elliott, 2007; Thurlow, 2007; Niebling & Elliott, 2005). Pitoniak and Royer (2001) describe accommodations as customizing the materials and testing conditions based on the examinee's needs. These changes are often categorized in terms of setting, timing, scheduling, presentation, and response.

Accommodations are intended to mitigate or lessen the impact of their disability but should not undermine the validity of the resulting test score interpretations (Hollenbeck, 2002; Sireci & Pitoniak, 2007). In other words, the purpose of an accommodation is to increase access as well as the accuracy of the measurement, and thereby the validity of the inference made from a test score. As described by the *Standards for Educational and Psychological Testing* (1999), the accommodations serve to “minimize the impact of test taker attributes that are not relevant to the construct that is the primary focus of the assessment” (p. 101). According to Green and Sireci (1999), a construct is a characteristic, skill, or ability that cannot be directly observed and that is believed to account for differences among individuals.

Accommodations serve, in part, to mitigate the impact or interaction of the disability with the content or skills assessed, by allowing the student to complete the test without the confounding influence of their disability (Hollenbeck, 2002). This type of interaction is unintended and introduces what Messick (1989) called construct-irrelevant variance, a serious threat to validity. Construct-irrelevant variance involves the degree to which factors that are not germane to the construct the test is intended to measure are in

fact reflected in the test score. Such variance can be caused by the interactions of the student's disability with either the administration protocols and/or the actual content measured on the test. Testing accommodations should neutralize or remove construct-irrelevant variance caused by the disability (Chiu & Pearson, 1999; Geisinger, 1994; Sireci & Pitoniak, 2007; Sireci, Scarpati, & Li, 2005; Thurlow, 2007). Described by Chiu and Pearson (1999), accommodations can be "thought of as a corrective lens" for potential score distortions caused by the student's disability and not their lack of knowledge and/or skill (p. 4).

Test developers and users have a responsibility to establish the utility and appropriateness of accommodations. Standard 10.1 in the *Standards for Educational and Psychological Testing* (1999) stipulates

In testing individuals with disabilities, test developers, test administrators, and test users should take steps to ensure that the test score inferences accurately reflect the intended construct rather than any disability and their associated characteristics extraneous to the intent of the measurement (p. 106).

This requirement is not easily achieved. Researchers are faced with the challenge of quantifying the effects of alterations in an effort to determine whether proposed accommodations fulfill their purpose of increasing accessibility while maintaining the technical quality of the assessment.

Historically, empirical studies of test administration alterations have focused largely on establishing the utility of accommodations for students with disabilities. This involves establishing the benefit of the accommodation for the targeted students, those with disabilities, as opposed to providing an advantage to those without disabilities, often

referred to as differential boost or the interaction hypothesis (Fuchs & Fuchs, 1999; Hollenbeck, 2005, 2002; Phillips, 1994; Sireci, Scarpati, & Li, 2005). The interaction hypothesis stipulates that when test administration accommodations are matched correctly based on student need, test performance will improve compared to the scores obtained under standard conditions. Furthermore, and importantly, students without disabilities will not benefit from the accommodation (Sireci, Scarpati, & Li, 2005). This protocol typically examines the interaction between student group membership (students with disability / students without disabilities) and test condition (accommodated / standard).

As originally conceptualized, the interaction hypothesis was said to be met if the performance of students with disabilities improved under the accommodated condition when compared to their performance under standard conditions *and* the performance of students without disabilities remained consistent under both conditions (accommodated and standard). Sireci, Scarprpati, & Li (2005) noted in their extensive literature review that often students without disabilities benefited from the accommodation as well, but not to the extent students with disabilities did. As a result, they suggested a revision, allowing gains for both groups of students but stipulated that gains for students with disabilities must be higher. According to Fuchs et al. (2000), differential boost requires students with disabilities must receive a significant test score increase compared to their nondisabled peers under the accommodated condition.

Read-aloud alteration. One controversial alteration is the read-aloud which permits students to have a test's entire text (both passages and questions) read aloud to facilitate their access. This alteration is typically considered for use by students who

have disabilities that impact their language processing or reading skill. Use of the read-aloud or oral administration is considered controversial when used on tests that measure reading comprehension because it is believed to encroach on the construct measured. Some advocates, however, argue that there are students whose disabilities prohibit their ability to access text and necessitate the need for the read-aloud or oral administration. They argue that unless decoding and fluency skills are explicitly contained within the test's constructs, often developed from state content standards, oral administration may be appropriate for some students (Laitusis, 2008, 2007).

In reporting the work of the National Accessible Reading Assessment Projects (NARAP), Thurlow (2010) suggests it is important to separate the individual components of reading into separate measures. She maintains that reading proficiency must include both comprehension and foundations skills such as decoding and fluency. No single component, such as phonemic knowledge, fluency, or comprehension, accounts for overall reading proficiency. She asserts, "students rely on their component proficiencies and the use of compensations to enable them to achieve overall proficiency in understanding a given text" (p. 124). By separating the components of decoding and fluency from comprehension, allowance of the read-aloud alteration may be permissible if the construct of interest for a test measuring comprehension is how well the student understands text. She cites college entrance exams such as the ACT and SAT as examples of testing programs that have defined the construct in this manner and allow the read-aloud as an accommodation for students with disabilities. Ensuring the construct of interest is clearly articulated is a critical step for any testing program (Downing & Haladyna, 2004) as is clearly articulating the interpretations that can and cannot be made

from the resulting test score.

Several studies have analyzed the impact of the read-aloud alteration on the reading test scores of students with and without disabilities. Findings have been mixed, with some studies finding evidence of differential boost for students with disabilities, while other studies did not. Meloy et al. (2002) investigated the impact of the read-aloud alteration on four subtests, including the reading comprehension subtest, of the Iowa Test of Basic Skills (ITBS), a norm-referenced test. The study included both students with and without disabilities but did not use a repeated-measures design, which Latusis (2007) considers to be the gold standard for research studies investigating accommodations. Rather, students were randomly assigned to a condition (standard or read-aloud administration). Both students with and without disabilities benefited from the read-aloud alternation. Although students with disabilities benefited more, the interaction between condition and group membership was not significant.

McKevitt and Elliott (2003) investigated the effects of a read-aloud administration on a norm-referenced test with eighth grade students. Both students with and without disabilities took the test under two conditions (standard and read-aloud). The read-aloud consisted of an audiotape recording of the test's content. Results were not significant and no interaction or differential boost was detected for the read-aloud condition. Neither group of students performed better under the accommodated condition, although students without disabilities performed better than their peers with disabilities (which was true for standard condition as well).

Crawford and Tindal (2004) examined the effects of the read-aloud alteration on a reading comprehension test with fourth and fifth grade students. In this study, the read-

aloud was administered through a video. A repeated measures design was employed and included both students with and without disabilities. Both groups benefited from the alteration; however, a significant interaction effect for students with disabilities was found.

In an alteration of the standard read-aloud administration in which all text is read to students, Fletcher et al. (2006) examined the impact of reading proper nouns and the item question to students with and without dyslexia. A repeated measures design was also employed and yielded a significant interaction between administration type and student group membership. Students with poor decoding skills benefited the most from the read-aloud alteration as operationalized in this study.

In a study involving both fourth and eighth graders, Laitusis (2010) found evidence of a differential boost in favor of students with reading-based learning disabilities. Although both groups of students, those with and without disabilities, benefited from the read-aloud condition, students with disabilities benefited more. This finding held even after controlling for reading fluency and ceiling effects (a concern for students without disabilities).

Randall and Engelhard (2010), using the same data set under consideration for this proposed study, found support for differential boost for fourth grade students, but not seventh grade students. In this repeated measures design, the performance of fourth grade students with disabilities improved with the read-aloud condition compared to their performance on the test during a standard administration. In grade 7, both student groups benefited.

Although results of studies investigating the read-aloud administration have been mixed, establishing the benefit of a test administration alteration for the targeted students (as opposed to providing an advantage to all students) is an important step when investigating the appropriateness of any alteration. Because reading is considered a fundamental gateway skill, it is imperative that use of read-aloud not be assigned to students in a wholesale fashion on a test designed to measure reading comprehension. If such a controversial alteration is to be allowed, it stands to reason that clear guidance must be established to ensure only students with demonstrated need are eligible. Such guidance should be grounded in empirical evidence.

The goal of the present study was to examine the effects of the read-aloud alternation on a test designed to measure reading comprehension. As discussed previously, student disabilities are often unique, with the same disability manifesting itself differently within individuals. Establishing a sufficient sample of students with similar needs is not easy due to the fact that for several disability categories the numbers are small, which limits the types of analyses that can be conducted (Pitoniak & Royer, 2001; Geisinger, 1994). Furthermore, there is significant variability in both the identification of students with disabilities and the assignment of disability categories (Barton, 2007). As a result, two students may be classified under the same disability category, yet their needs may be very different. Many studies have sampled students with disabilities as a group, rather than sampling the students who have exhibited a specific need for the alteration under consideration. Other studies have sampled a specific category, such as students with learning disabilities.

Malouf (2005) suggests that studies that group students with like disability

categories are less informative than studies that focus on a specific accommodation and define the student group based on common characteristics the accommodation is intended to mitigate. At a recent technical assistance meeting hosted by the US Department of Education, Thurlow (2011) advocated the need for improved selection of students for participation in research studies, ensuring studies focus on students who demonstrate need and would potentially benefit from the accommodation of interest. As Tindal et al. (1998) summarize, “to provide the most convincing empirical support for an accommodation, students with a specific need have to be compared to others without such a need who are otherwise comparable in achievement” (p. 442).

This study is designed to investigate if students with disabilities with certain characteristics benefit from the use of the read-aloud alteration. The study sought to examine several potential moderator variables to illuminate the characteristics of students with disabilities who benefited. It was hypothesized that the moderator variables would help identify which students are likely to need and benefit from the read-aloud alteration. Specially, five research questions were addressed:

1. For third grade students with disabilities receiving the read-aloud alteration, do students with below-average reading skills achieve greater gain scores (or a differential boost) in reading performance compared to students, with and without disabilities, with average to above average reading skills? (It was hypothesized that students with disabilities who had below average reading skills would benefit more from the read-aloud alteration.)
2. Do students with disabilities whose teachers rated their degree of disability moderate or severe differentially benefit from the read-aloud alteration when compared to peers

- rated as having a mild disability? (It was hypothesized that students with more severe disabilities would benefit more from the read-aloud alteration.)
3. Do students with disabilities whose Individualized Educational Programs (IEP) include long-term objectives that address decoding skills differentially benefit from the read-aloud alteration compared to peers whose IEPs do not include long-term objectives for decoding? (It was hypothesized that students with decoding objectives would benefit more from the read-aloud alteration.)
 4. Do students with disabilities who spend 50 percent or less of their instruction time in the general education classroom differentially benefit from the read-aloud alteration compared to peers who spend more than 50 percent their time in the general education classroom? (It was hypothesized that students spending less than 50 percent of their time in the general education classroom would benefit more from the read-aloud alteration.)
 5. Do students with disabilities whose receive below grade-level reading instruction differentially benefit from the read-aloud alteration compared to peers who receive reading instruction on grade level? (It was hypothesized that students who received below-grade level reading instruction would benefit more from the read-aloud alteration.)

Extant data collected as part of a larger research study designed by the student and conducted by the Georgia Department of Education in 2006 under the student's direction as the Assistant Director of Testing was used. An external criterion, the reading comprehension subtest of the Iowa Test of Basic Skills (ITBS), was used to classify student reading skill into three levels – below average, average, and above average.

Additional aspects about special education students' educational programs were also included in the analyses. These factors were derived from a survey completed by a teacher for each special education student who participated in the study.

Methods

Participants

Participants consisted of fourth-grade students enrolled in public schools throughout the state of Georgia during the 2005-2006 school year. A stratified-random sample of schools was selected to mirror enrollment demographics of the state's third grade students during the 2004-2005 school year. Schools were stratified into three levels based on the proportion of students receiving free or reduced price lunch (as a proxy for socio-economic status) and randomly selected for participation. To ensure representativeness, additional demographics were reviewed including gender, race/ethnicity, disability category classification (i.e., specific learning disability, etc.) and achievement as measured by the grade 3 state-mandated reading assessment administered in the spring of 2005. In other words, the sample was selected to be representative of the grade 3 state student population although the sample consisted of fourth-grade students.

According to the Georgia Department of Education website, approximately 1.5 million students were enrolled in the public school system (K – 12) during the 2004-2005 school year. Of these, approximately 12.2% received special education services and a total of 117,298 students were enrolled in grade three. For the study sample, both students with and without disabilities were included, although students with disabilities were oversampled to ensure approximately equal numbers. The final selected sample included eight districts (out of 183) and 129 schools (out of approximately 2100)

representing all geographical regions of the state, including urban, suburban, and rural areas. A total of 754 fourth-grade students participated in the study. The Spring 2005 grade 3 results, utilized as pretest measures to allow for within subject comparisons, were drawn from the state-mandated administration to all third grade students in the state under standard, operational conditions. The Spring 2006 study data file was matched to the Spring 2005 grade 3 Reading CRCT state data file. A match rate of 88% was achieved, resulting in 664 students with a complete record (i.e., both 2005 and 2006 scores). Of these, 316 (48%) were students with disabilities and 348 (52%) were students without disabilities. Table 4 provides the demographic breakdown by student disability status and assigned test condition. Finally, Table 5 provides the number of students with disabilities sampled by IDEA disability category.

An external criterion measure, the Reading Comprehension subtest of the Iowa Test of Basic Skills (ITBS), was used to classify the level of student reading skill as below average or average/above average. All third grade students participated in the complete battery of the ITBS as part of the state-mandated testing program under standard, operational conditions during the 2004-2005 school year. A match rate of 73% was achieved, resulting in 491 complete records with all three test records (ITBS, CRCT Spring 2005, and CRCT Spring 2006). Of these, 230 (47%) were students with disabilities and 261(53%) were students without disabilities. Table 6 provides the demographic breakdown by student disability status and assigned test condition for the ITBS-matched sample.

Table 4

Demographic Characteristics of Fourth-Grade Students by Disability Status and Test Condition

Characteristic	Test condition			
	Read Aloud		Standard	
	Students without Disabilities <i>n</i> = 146	Students with Disabilities <i>n</i> = 145	Students without Disabilities <i>n</i> = 202	Students with Disabilities <i>n</i> = 169
Gender				
Male	75	98	104	112
Female	71	48	97	54
Race/Ethnicity				
Asian, Pacific Islander	6	6	10	9
Black, Non-Hispanic	62	59	40	50
Hispanic	5	6	21	15
American Indian/ Alaskan Native	0	0	0	0
White, non-Hispanic	63	72	119	86
Multiracial	10	2	12	9

Table 5

Number of Students with Disabilities in Each IDEA Category

Disability	<i>N</i>
Specific Learning Disability	121
Speech Language Impairment	72
Intellectual Disabilities	28
Hearing Impaired	9
Visual Impairment	0
Emotional Disturbance	27
Orthopedic Impairment	3
Other Health Impairment	50
Autism	8
Traumatic Brain Injury	0
Multiple Disabilities	0
Total	318

Table 6

Demographic Characteristics of Fourth-Grade Students by Disability Status and Test Condition for the ITBS-Matched Sample

Characteristic	Test Condition			
	Read Aloud		Standard	
	Students without Disabilities <i>n</i> = 103	Students with Disabilities <i>n</i> = 101	Students without Disabilities <i>n</i> = 158	Students with Disabilities <i>n</i> = 129
Gender				
Male	52	65	82	88
Female	51	35	75	38
Race/Ethnicity				
Asian, Pacific Islander	5	6	7	6
Black, Non-Hispanic	36	30	31	36
Hispanic	5	4	14	11
American Indian/ Alaskan Native	0	0	0	0
White, non-Hispanic	48	59	98	68
Multiracial	9	1	8	8

Procedure

Superintendents, test coordinators, and special education directors in the selected districts were notified that certain schools within their district were selected to participate in the state's study by the Georgia Department of Education's Deputy State Superintendent of Curriculum and Instruction (see Appendix A). Selected schools were

randomly assigned into one of two test conditions: a) read-aloud administration or b) standard administration. Each school was told the number of fourth-grade students with disabilities to test based on the number of grade 3 students tested in the spring of 2005. All students with disabilities in the school were selected to participate with the exception of those students who required either a Braille or large-print edition of the test or those students with significant cognitive disabilities who met the participation criteria for the state's alternate assessment based on alternate achievement standards (AA-AAS). Schools were instructed to randomly identify an equal number of students without disabilities to participate as well; districts were provided examples of how to randomly select students without disabilities. All students, both students with and without disabilities, in the school were tested under the assigned condition.

Each participating school received a brief summary that included the rationale, purpose, and logistics for the study including an introduction to administration procedures specific to each test condition (see Appendix B). Students with disabilities were allowed to use other needed accommodations, as identified in their IEPs, in addition to the assigned test condition, provided those accommodations resulted in a standard administration. Schools were instructed not to provide accommodations that resulted in a non-standard administration. The *Examiner's Manual* that accompanied the test materials provided a list of accommodations, including those that were considered non-standard.

In February 2006, district test coordinators participated in pre-administration webinar that outlined the study rationale, sample selection, assigned test conditions, and administration procedures. The sampled fourth-grade students were administered the

third-grade Reading CRCT during a 2-week window in March 2006. The same third-grade test form administered operationally during the Spring of 2005 was used. Standard administration and test security procedures were followed. Schools assigned the read-aloud condition were instructed to read the entire test to students, including the reading passages, test questions, and response options. Examiners were instructed to read the test verbatim and to read at a natural pace.

In addition to administering the test under the condition assigned, schools were asked to have the educator most knowledgeable about the student complete a 20-question survey for each special education student participating in the study regardless of the test condition assigned to the school. The survey was designed to capture pertinent information about the student's disability, educational program, and instructional needs.

Instruments

The Reading CRCT is a 50-item test designed to measure student acquisition of the knowledge and skills inherent in the state's content standards. The test consists of 40 operational items which contribute to the student's score and ten field-test items which do not contribute to the student's score, and is administered in two 60 – 70 minute sections. The test was developed in 2000 by the Georgia Department of Education as a criterion-referenced test for the purposes of measuring student mastery and gauging the quality of instructional services provided throughout the state. State law mandated that students in grades 1 – 8 be assessed annually in reading, English/language arts, and mathematics. Students in grades 3 – 8 are also assessed in science and social studies.

The grade 3 Reading CRCT is comprised of four domains, groupings of items measuring standards with similar content and skills, according to the CRCT Content

Descriptions for Reading (Georgia Department of Education, 2001). The four domains include a) Reading for Vocabulary Improvement; b) Reading for Locating and Recalling Information; c) Reading for Meaning; and d) Reading for Critical Analysis. The Vocabulary Improvement domain consists of nine items measuring the students' ability to apply word attack and recognition skills (such as context clues, root words, suffixes, and prefixes), use word knowledge (such as antonyms, synonyms, word order and syntax), and interpret words and phrases. Given that vocabulary serves as a primary building block for comprehension, words assessed are taken from the passages within the test. The Locating and Recalling Information domain consists of 10 items and assesses the students' skill in recognizing and recalling information from a variety of texts including important and supporting details. The Reading for Meaning domain has 12 items that measure the students' skill at identifying literary forms (such as fiction, nonfiction, poetry), the purpose of a text, characters and their traits, text features such as sequence and organization, as well as recognizing the explicit main idea of a passage. Finally, the Critical Analysis domains consists of nine items assessing students' skill in making predictions and generalizations from text, comparing and contrasting, and drawing conclusions or making inferences, as well as recognizing the implicit main idea of a passage. The test was developed to align to the state content standards enacted at that time, the Quality Core Curriculum. According to the Spring 2005 CRCT Technical Report the reliability of the grade 3 Reading CRCT was .88 (Cronbach's alpha, a measure of internal consistency). The total number of items correctly answered results in the CRCT reading raw score, which can range from 0 to 40.

The Iowa Test of Basic Skills (ITBS) is nationally norm-referenced test administered to Georgia public school students in grades 3, 5, and 8 as part of the state-mandated testing program. Third grade students took Form A/Level 9 of the Complete Battery of the ITBS in Fall 2004. According to the *Guide to Research and Development* (Hoover et. al., 2003), the purpose of the test is to “measure growth in fundamental areas of school achievement” (p. 1). The Reading Comprehension subtest is comprised of 37 questions administered in two sections of 25 and 30 minutes each. ITBS items are designed to measure students’ Factual Understanding (including stated information and words in context), Inference and Interpretation (including drawing conclusions, inferring character traits, feelings, and motives, interpreting information in new contexts), and Analysis of Generalization (determining main idea, identifying the author’s purpose or viewpoint, and analyzing a passage’s style or structure). The reliability of the Reading Comprehension subtest (Fall 2000 norms) is reported as .896 (Kuder-Richardson Formula 20; also a measure of internal consistency) (Hoover et. al., 2003).

Schools were asked to complete a Special Education Survey for each student with a disability participating in the study. The 20-item survey was adapted, with permission, from one used by the National Center for Educational Statistics (NCES) for the National Assessment of Educational Progress (NEAP). Questions included, but were not limited to, the areas of the student’s disability, the content areas in which the student received special education services, the teacher’s rating of the severity of the disability, the presence of long-term objectives in the student’s IEP in reading and mathematics, the proportion of academic class time spent in the general education classroom, the grade

level of instruction provided the student, and the manner in which the student participated in assessments. A copy of the survey is included in Appendix C.

Data Analysis

Student test data were analyzed using a series of factorial ANOVAs. This approach allows for the evaluation of the effectiveness of one or more experimental conditions, such as the two test conditions (read aloud / standard administration) with a control group (students with and without disabilities took the test under the standard condition). Reading scores from the Spring 2005 administration served as the baseline or pretest measure and were compared to scores obtained during the second administration in March 2006. Gain scores were calculated by subtracting the Spring 2005 raw score (number correct) from the Spring 2006 raw score. As outlined earlier, the same form of the test was administered allowing the raw score to be utilized in this manner.

Prior to conducting the ANOVA, data were screened and all assumptions, such as homogeneity of variance and normality, evaluated. The dependent variable for all research questions is the students' Reading CRCT gain score. For research question 1, there are three independent variables, a) testing condition (i.e., read-aloud and standard administration), b) disability status (students with and without disabilities), and c) reading skill of the student. Students' reading skills was classified as below average or average/above average using the normal curve equivalency (NCE) score from the Fall 2004 third-grade Reading Comprehension subtest of the ITBS. According to the Interpretative Guide for School Administrators (Hoover et al., 2003), NCEs are normalized standard scores with a mean of 50 and a standard deviation of 21.06. Similar to percentiles, they range from 1 to 99 but unlike percentiles, they are equal intervals and

can be averaged. NCEs of 1 – 34 are considered below average, 35 – 65 are considered average, and 66 – 99 are considered above average. To evaluate research question 1, the gain scores of students with and without disabilities who took the Reading CRCT under the read-aloud or standard administration condition was compared to their peers based on their ITBS NCE classification.

Research questions 3 – 5 focus specifically on students with disabilities, a subset of the participants. Using the Special Education Survey, students were classified according to the survey results and the gain scores compared for those who took the March 2006 administration under the read aloud or standard administration conditions. For research question 2, students were classified into one of two categories based on the teacher's rating of the severity of the student's disability (mild or moderate / profound / severe). For research question 3, students were classified into one of two categories based on whether their IEPs included long-term objectives address the student's ability to decode simple printed material. For research question 4, students were classified into one of two categories describing the amount of academic class time sent in the mainstream/general education classroom (50% or less of time or more than 50% of time). For research question 5, students were classified according to whether their teacher reported providing below grade-level reading instruction or above grade level.

Within this study, it was hypothesized that some of the independent variables will act as moderators. Moderator variables are hypothesized to influence the strength or direction of the relationship between two other independent variables (Baron & Kenny, 1986). The Reading Comprehension NCE for the ITBS is hypothesized to be a moderator variable, with the hypothesis that students who received a below average NCE

will benefit more from the read-aloud condition than their peers who received an average or above average NCE. Additionally, the survey classifications will serve as moderators; in these cases, the moderator variables are hypothesized to help better identify students with disabilities who are in need of and would benefit from the read-aloud alteration. For instance, it is hypothesized that students whose disabilities are classified as moderate or severe by their teacher would benefit more from the read-aloud administration than their peers who are classified as having a mild disability.

The statistical tests of interest from the factorial ANOVA was the interaction effect. When interactions were detected, post hoc tests were conducted in order to determine where those difference occurred and identify potential interactions among student group membership (students with or without disabilities) and test condition (read aloud or standard), as well as among the moderator variable classifications. Because multiple ANOVAs were conducted, a Bonferroni adjustment was made, and a more conservative alpha level was used ($p < .01$).

Results

The data were screened for outliers and assumptions were examined. A decision rule was made that students who exhibited a decrease of 20 raw score points or more on the posttest ($N = 22$) lacked motivation. These students were excluded from the analyses. The reliability of the test was examined by condition and disability status. The posttest alpha coefficients were .89 or above indicating that the overall internal consistency of the test remained strong regardless of test administration condition (read aloud or standard) or student group (students with disabilities or students without disabilities). The assumptions of homogeneity of variance and normality were examined before conducting

the major analyses. All the assumptions appeared tenable resulting in unbiased results for all research questions.

Disability Status and Level of Reading Skill

Research Question #1. Do third-grade students with disabilities who have below-average reading skills differentially benefit from the read-aloud compared to students, with and without disabilities, with average to above average reading skills?

A three-way ANOVA was used to examine research question 1. The independent variables were a) testing condition, b) disability level, and c) reading skill. The dependent variable was the raw score gain from pretest to posttest. The means, standard deviations, and sample sizes are reported in Table 7. The mean raw score gain ranged from 0.23 for students with disabilities who had below average reading skills according to the ITBS and were assessed under the standard condition, to 8.30 for students without disabilities who had below average reading skills who were provided the read-aloud condition. Students with disabilities who were classified as below average and provided the read-aloud condition had a mean gain score of 7.58.

Table 7

Means, Standard Deviations, Sample Sizes of the Raw Score Gain by Disability Status, Test Condition and Level of Reading Skill

Disability Status	Reading Skill	<i>Standard Condition</i>			<i>Read-aloud Condition</i>		
		<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
SWOD	Below	2.50	5.53	8	8.30	7.17	10
	Average	0.85	5.21	75	2.95	6.39	56
	Above	1.27	3.06	64	1.92	3.46	26
SWD	Below	0.23	7.17	30	7.58	7.48	33
	Average	0.69	6.25	78	2.52	5.83	48
	Above	1.16	5.88	19	1.81	3.53	16

Note: SWOD = Students without Disabilities; SWD = Students with Disabilities.

The results of the three-way ANOVA are reported in Table 8. There was a statistically significant main effect for testing condition ($F_{(1, 451)}=27.00, p<.001$) and reading skill ($F_{(2, 451)}=6.89, p=.001$). Students in the read-aloud condition had a higher raw score gain ($M=3.69, SD=6.33$) than students in the standard condition ($M=.91, SD=5.41$). Students in the below average reading group had higher gains ($M=4.44, SD=7.90$) than students in the average reading group ($M=1.57, SD=5.97$) and students in the above average reading group ($M=1.46, SD=3.17$). There was a statistically significant testing condition by reading skill interaction effect ($F_{(1, 451)}=7.42, p=.001$). The interaction can be seen in Figure 1. Across all reading skill levels, students in the read-aloud condition had higher gain scores than those in the standard condition. The greatest differences were seen in the below average reading skill groups. Students in the

below average reading skill who were given the read-aloud condition had higher gains ($M=7.74$, $SD=7.33$) than those students in the below average reading skill in the standard testing condition ($M=.71$, $SD=6.85$). The *a priori* hypothesis suggested students with a disability who received a read-aloud and were in the below average reading group would have greater gains than other conditions. While these students demonstrated large gains ($M=7.58$, $SD=7.48$), so did the students who did not have a disability who were in the below average skill group and received a read-aloud accommodation ($M=8.30$, $SD=7.17$).

Table 8

Three-way ANOVA Results for Testing Condition by Disability Status and Level of Reading Skill

Source	df	F	Sig	Partial Eta
Testing Condition	1	27.00	<.01	.06
Disability Status	1	.03	.87	<.01
Reading Skill	2	6.89	<.01	.03
Testing*Disability	1	1.32	.25	<.01
Testing*Reading Skill	2	7.42	<.01	.03
Disability*Reading Skill	2	.28	.76	<.01
Testing*Disability*Reading Skill	2	.14	.87	<.01

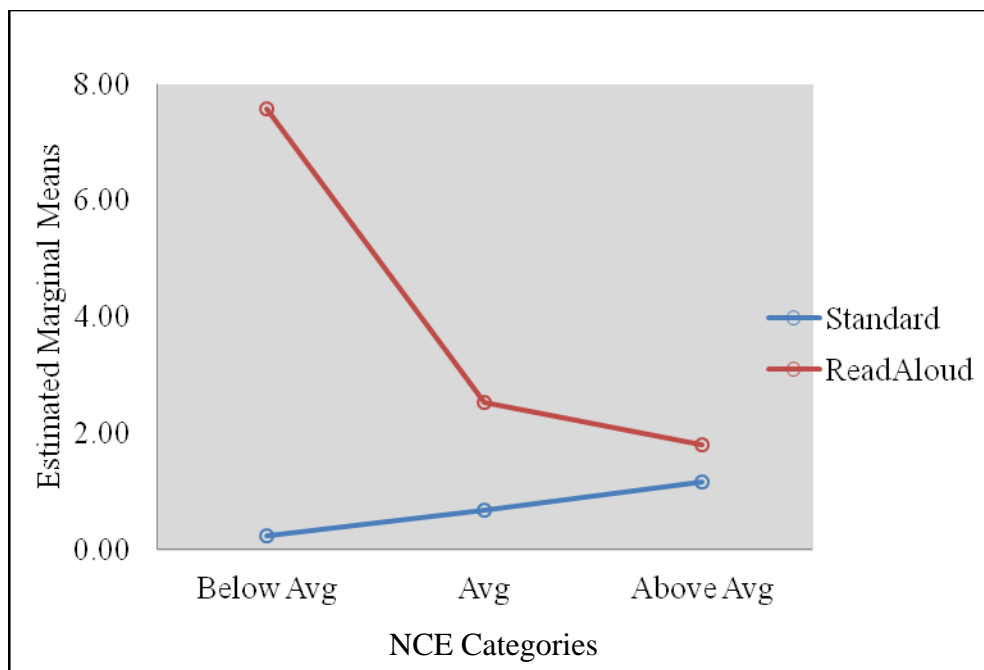


Figure 1. The interaction of testing condition by level of reading skill for students with disabilities.

The effect size measures (Cohen's d) for differences between the standard condition and the read-aloud condition for students with and without disabilities are presented in Table 9. For both students with and without disabilities, there were very large differences between the standard and read-aloud conditions for students with below reading skill, 1.03 and 1.05 respectively. While the read-aloud conditions had higher gain scores than the standard condition in the average and above average reading skill categories, the size of the effects were much smaller. Readers should be cautious interpreting the effects due to a ceiling effect (i.e., higher ability reading students tended to score at the highest level). Approximately 57% of students classified as having above average reading skill had a raw score of 35 points or higher, out of a possible 40, on the pretest. As comparison, only 6% of students classified as having average reading skills and none of the student classified as having below average skills had a raw score of 35 or higher on the pretest.

Table 9

Effect sizes for Reading Skill by Disability Status

Reading Skill	SWD	SWOD
	<u>d</u>	<u>d</u>
Below	1.03	1.05
Average	.29	.40
Above Average	.11	.21

Note: SWOD = Students without Disabilities; SWD = Students with Disabilities.

Degree of Disability

Research Question #2. Do students with disabilities whose teachers rated their degree of disability moderate or severe differentially benefit from the read-aloud compared to peers rated as having a mild disability?

A two-way ANOVA was used to examine research question 2. The independent variables were a) testing condition and b) degree of disability. The dependent variable was the raw score gain from pretest to posttest. The means, standard deviations, and sample sizes are reported in Table 10. Students whose teachers rated their degree of disability as mild had higher gain scores than those students rated as having moderate or severe/profound disabilities, regardless of whether they were assessed under the standard or read-aloud condition.

Table 10

Means, Standard Deviations, Sample Sizes of the Raw Score Gain by Test Condition and Degree of Disability

Degree of Disability	<i>Standard Condition</i>			<i>Read-aloud Condition</i>		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
Mod/Prof	.44	7.40	54	3.19	7.68	36
Mild	.60	5.08	77	3.74	7.78	89

Note: Mod/Prof = Moderate/Profound.

The results of the two-way ANOVA are reported in Table 11. There was a statistically significant main effect for testing condition ($F_{(1, 252)}=10.13, p<.001$). Students administered the read-aloud condition had a higher raw score gain ($M=3.58, SD=7.73$) than students in the standard condition ($M=.53, SD=6.12$). The *a priori* hypothesis suggested students with a disability who received the read-aloud condition and were rated by their teacher as having a moderate or profound disability would have greater gains than those rated as having a mild disability. Students who were rated as having a mild disability and were assessed under the read-aloud condition posted larger gains ($M=3.74, SD=7.78$) than their peers rated as having moderate or profound disabilities ($M=3.19, SD=7.68$) tested under the same condition. There was no significant interaction, however, between testing condition and degree of disability ($F_{(1, 252)}=.045, p=.832$).

Table 11

Two-way ANOVA Results for Testing Condition by Degree of Disability

Source	df	F	Sig	Partial Eta
Testing Condition	1	10.13	<0.01	.04
Degree of Disability	1	.14	.71	<.01
Testing*Degree	1	.05	.83	<.01

IEP Objective for Decoding

Research Question #3. Do students with disabilities whose Individualized Education Programs (IEP) include long-term objectives that address decoding skills differentially benefit from the read-aloud compared to peers whose IEPs do not include long-term objectives for decoding?

A two-way ANOVA was used to examine research question 2. The independent variables were a) testing condition and b) presence of an IEP long-term objective addressing decoding. The dependent variable was the raw score gain from pretest to posttest. The means, standard deviations, and sample sizes are reported in Table 12. According to teachers, the vast majority of students with disabilities ($N=224$) included in the study had decoding addressed in the IEP as a long-term objective.

The results of the two-way ANOVA are reported in Table 13. No significant main effect for testing condition ($F_{(1, 248)}=2.39, p=.123$) was found; nor was a significant interaction detected between testing condition and presence of a decoding objective in the IEP ($F_{(1, 248)}=.709, p=.401$).

Table 12

Means, Standard Deviations, Sample Sizes of the Raw Score Gain by Test Condition and Inclusion of Decoding Objectives in IEP

IEP	<i>Standard Condition</i>			<i>Read-aloud Condition</i>		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
Decode	.21	6.11	112	3.60	7.94	112
No Decode	2.21	6.99	14	3.21	6.82	14

Table 13

Two-way ANOVA Results for Testing Condition by Inclusion of Decoding Objectives in IEP

Source	df	F	Sig	Partial Eta
Testing Condition	1	2.40	.12	.01
IEP Decoding Objective	1	.33	.57	<.01
Testing*IEP	1	.71	.40	<.01

Time Spent In General Education Classroom

Research Question #4. Do students with disabilities who spend 50 percent or less of their instructional time in the general education classroom differentially benefit from the read-aloud compared to peers who spend more than 50 percent of their time in the general education classroom?

A two-way ANOVA was used to examine research question 4. The independent variables were a) testing condition and b) time in general education classroom. The dependent variable was the raw score gain from pretest to posttest. The means, standard deviations, and sample sizes are reported in Table 14. Students who were reported as

spending less than 50% of their time in the general education classroom had lower raw scores on the posttest ($M=-.31$, $SD=7.33$).

Table 14

Means, Standard Deviations, Sample Sizes of the Raw Score Gain by Test Condition and Percent of Time Spent in General Education Classroom

Time in General Education Classroom	<u>Standard Condition</u>			<u>Read-aloud Condition</u>		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
< 50%	-2.33	5.88	42	1.46	8.05	48
>50%	1.83	5.77	90	4.71	7.37	80

The results of the two-way ANOVA are reported in Table 15. There was a statistically significant main effect for testing condition ($F_{(1, 256)}=12.64$, $p<.001$). Students with disabilities administered the read-aloud condition had a higher raw score gain ($M=3.49$, $SD=7.76$) than their peers in the standard condition ($M=.508$, $SD=6.10$). Additionally, a statistically significant main effect for percent of time spent in the general education classroom was found ($F_{(1, 256)}=17.53$, $p<.001$). Students with disabilities who were reported as spending less than 50% of their time in a general education classroom had a lower gain score mean ($M=-.311$, $SD=7.33$) than their peers who were reported as spending more than 50% of their time in a general education classroom ($M=3.19$, $SD=6.71$). No interaction between testing condition and percent of time spent in the general education classroom was detected ($F_{(1, 256)}=0.27$, $p=.606$).

Table 15

Two-way ANOVA Results for Testing Condition by Time Spent in General Education Classroom

Source	df	F	Sig	Partial Eta
Testing Condition	1	12.64	<.01	.05
General Education Instruction	1	17.53	<.01	.06
Testing*General Education	1	.27	.61	<.01

Grade Level of Reading Instruction

Research Question #5. Do students with disabilities who receive below grade level reading instruction differentially benefit from the read-aloud compared to peers who receive reading instruction on grade level?

A two-way ANOVA was used to examine research question 5. The independent variables were a) testing condition and b) grade level of reading instruction. The dependent variable was the raw score gain from pretest to posttest. The means, standard deviations, and sample sizes are reported in Table 16. Students who received below grade level instruction and were tested under the standard condition score lower on the posttest than the pretest ($M=-1.48$, $SD=5.97$). Students receiving instruction on grade level and provided the read-aloud alteration posted the highest gain scores ($M=4.30$, $SD=6.56$).

Table 16

Means, Standard Deviations, Sample Sizes of the Raw Score Gain by Test Condition and Grade Level of Reading Instruction

Instruction	<i>Standard Condition</i>			<i>Read-aloud Condition</i>		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
On Grade	1.83	5.98	77	4.30	6.56	63
Below Grade	-1.48	5.97	52	2.78	8.94	64

The results of the two-way ANOVA are reported in Table 17. There was a statistically significant main effect for testing condition ($F_{(1, 252)}=12.17, p<.001$). Students with disabilities administered the read-aloud condition had a higher raw score gain ($M=3.54, SD=7.86$) than students in the standard condition ($M=.496, SD=6.17$). A statistically significant main effect was also detected for grade level of instruction ($F_{(1, 252)}=7.48, p<.001$), with students who were reported as receiving instruction on grade level having a higher raw gain score ($M=2.94, SD=6.35$) than those reported as receiving instruction below grade level ($M=0.87, SD=8.01$). No significant interaction between testing condition and grade level of instruction was found ($F_{(1, 252)}=1.04, p=.31$),

Table 17

Two-way ANOVA Results for Testing Condition by Grade Level of Reading Instruction

Source	df	F	Sig	Partial Eta
Testing Condition	1	12.17	<.01	.05
Grade Level Instruction	1	7.48	.01	.03
Testing*Grade Level	1	1.04	.31	<.01

Discussion

The purpose of this study was to examine the effects of a controversial test administration alteration, the read-aloud alteration, in which both reading passages and questions are read aloud to the student on a test designed to measure reading comprehension. For students whose disabilities impair their skills in decoding text and reading fluently, accessing text to demonstrate their comprehension can be significantly impeded. Thurlow (2010) suggests that no single component process, be it phonemic knowledge, decoding skills, reading fluency, or comprehension, accounts for overall reading proficiency. This study examined whether the test scores for students with disabilities with certain characteristics improved when the read-aloud alteration was provided.

In general, students assessed under the read-aloud condition achieved higher raw score gains on the Reading CRCT than those assessed under the standard condition regardless of their disability status (students with or without disabilities). No interactions were identified between the moderator variables studied and test condition (standard or read loud), with the exception of testing condition and reading skill (below average, average, or above average). Regardless of disability status, students who were provided the read-aloud alteration and were classified as having below-average reading skills on the norm-referenced ITBS had higher gain scores than their peers. Other moderator variables studied included degree of disability, presence of a decoding objective in the IEP, time spent in the general education classroom, and grade level of reading instruction.

This study was designed to address sampling concerns raised by several researchers who have studied the effect of test administration accommodations or

alterations (Barton, 2007; Geisinger, 1994; Malouf, 2005; Pitoniak & Royer, 2001; Thurlow, 2011; Tindal et al., 1998). Investigating the effects of any test administration alteration and the resulting impact on the meaning of the test score is a complicated endeavor. Given the heterogeneity of the population of students with disabilities, acquiring an adequate sample of students with similar needs to access both content and a test is difficult (Geisinger, 1994; Pitoniak & Royer, 2001). Samples using disability categories as selection criteria, such as learning disabled, can inadvertently limit the generalizations of the findings. Discrepancies in the identification and disability category classification techniques contribute to this problem (Barton, 2007). As previously discussed, how disabilities manifest themselves within individuals varies greatly. For example, learning disabilities can impact listening, thinking, speaking, reading, writing, spelling, and/or mathematical calculations (Pitoniak & Royer, 2001). And if, as Fuchs, Fuchs, & Capizzi (2005) assert, what might be a strength for one student identified as having a learning disability may be a weakness for another, using a disability category as a selection criterion is not likely to garner a homogeneous sample.

This study sought to address these concerns through oversampling the target population, students with disabilities, while keeping the proportion of the IDEA disability categories within the sample representative of the state. Whereas students with disabilities comprise approximately 12% of Georgia's public school student population, the sample selected for this study was approximately half students with disabilities and half students without disabilities. The large sample allowed for the investigation of potential moderator variables with the objective of identifying student and educational

program characteristics, regardless of disability category, that might illuminate which students might benefit from the read-aloud alteration.

While the study did not uncover significant interactions between the moderator variables studied and testing condition as hypothesized, it is an example of the type of research that should contribute to the efforts IEP teams undertake when making accommodation decisions. According to Hollenbeck (2005), IEP teams do not always have the knowledge and understanding necessary to make competent decisions, particularly when it comes to test administration accommodations. He found that teachers have difficulty differentiating between students who would benefit from an accommodation from those who would not, as well as predicting which accommodations would be helpful. He further asserts that the most significant threat to validity stemming from accommodation usage is inconsistent selection and implementation. He argues that, too often IEP teams “make decisions idiosyncratically and unsystematically” (p. 415). Other researchers concur, finding that educators tend to make decisions based on their perceptions about accommodations rather than guidance, policy, or differentiated need (Gilbertson Schulte, Elliott, and Kratochwill, 2000; Lang et al., 2005).

A quality study of test administration accommodations should consist of a rigorous experimental design that includes both an adequate sample size and random assignment of accommodation conditions to students with and without disabilities (Laitusis, 2007). A major strength of this study was the large and inclusive sample, as was the repeated measures design. A limitation of the study was the random assignment of the test condition (read aloud or standard) at the school level rather than at the individual student level. This was done to make the administration for the participating

schools and the logistics for carrying out such a large-scale study manageable. While this approach may have been less than desirable, the fact that the assignment of test condition at the school level was random reduces the threat to the conclusions drawn from the study. Another limitation of the study was the ceiling effect experienced by the students who were classified as reading above average.

In interpreting the results, it is important to consider that student scores from the study were compared to student's test scores derived under operational conditions in the spring of the previous year. Georgia law mandates that all students take the CRCT each spring and stipulates that students in grades 3 pass the Reading CRCT in order to be considered eligible for promotion to the fourth grade. Given the stakes involved, students were likely highly motivated during the Spring 2005 operational administration and less motivated during the study administration for which there were no consequences for students or schools. For this reason, students who had highly discrepant gain scores (≤ 20) were removed from data analysis.

The implications of this study are clear: use of the read-aloud alteration is not appropriate for all students and may provide an unfair advantage if assigned inappropriately to students. In general, students who received the read-aloud alteration benefited more than those who did not, regardless of their disability status (with or without a disability) or the presence of a long-term decoding objective in their IEP. Students with disabilities who were rated as having a mild disability, who spent more than 50 percent of their time in the general education classroom, and who received grade-level reading instruction tended to achieve higher gain scores than their peers. It was hypothesized that students with more severe disabilities, who spent less than 50 percent

of their time in the general education classroom, and who received reading instruction below their grade level would benefit more from the read-aloud condition. It was thought that the read-aloud condition would help these students compensate for their lower reading skills and allow greater access to the passage content so that the students could demonstrate a higher level of comprehension. This did not prove to be the case in this study. Perhaps the students who did benefit had more opportunity to learn and practice the reading skills important to demonstrating reading comprehension, given their increased time in the general education setting and exposure to grade-level instruction.

Without a doubt, the provision of a read-aloud alteration on a test designed to measure reading comprehension is controversial. Virtually all identified studies of the read-aloud alteration have found that all students, both with and without disabilities, benefit from the provision of the read-aloud (Crawford & Tindal, 2004; Laitusis, 2010; McKeivite & Elliott, 2003; Meloy et al., 2002; Randall & Englehard, 2010). And while some of these studies found a significant interaction in favor of students with disabilities, the current study did not. While results for the read-aloud alteration remain mixed, it is clear that the read aloud is not appropriate for all students and it remains to be seen if, indeed, it is appropriate for some students with disabilities. Further research is needed to better understand the characteristics associated with the disabilities that create barriers for students in accessing text for both instructional and assessment purposes. Opportunity to learn may be a key consideration given the historical lack of academic opportunities provided to students with disabilities.

Ultimately, this study is an example of the types of investigations that state departments of education can conduct to inform the assessment policies and practices of

their operational programs. This study utilized the overage of test materials (i.e., unused answer documents, test books, and examiner's manuals) from the spring operational administration and was relatively inexpensive to conduct. Due to a variety of resource constraints, such as financial, human, and practical, states rarely undertake special empirical investigations such as this. Given the equity issues surrounding access to educational opportunities for students with disabilities, making informed decisions about accommodations are essential to the integrity of the assessment. After all, ensuring accurate information about student achievement results so that appropriate and effective instructional decisions can be made is the paramount purpose of assessment.

References

- All Handicapped Children Act, 10, U.S.C. 1400 *et seq.* (1975).
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standard for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Americans with Disabilities Act of 1990, 42 U.S.C. § 12101 *et seq.*
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.
- Barton, K. E. (2007). Validity and accommodations: The journey toward accessible assessments. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 81-94). Arlington, VA: Council for Exceptional Children.
- Chiu, C. WT, & Pearson, P. D. (1999, June). *Synthesizing the effects of test accommodations for special education and limited English proficient students*. Paper presented at the National Conference on Large-Scale Assessment, Snowbird, UT.
- Cormier, D. C., Atلمان, J. R., Shyyan, V., & Thurlow, M. L. (2010). A summary of the research on the effects of test accommodations: 2007-2008 (Technical Report 56). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

- Crawford, L. & Tindal, G. (2004). Effects of a read-aloud accommodation on a standardized reading test. *Exceptionality, 12*(2), 89-106.
- Downing, S. M. & Haladyna, T. M (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement in Education, 10*(1), 61-82.
- Education for All Handicapped Children Act, 20 U.S.C. § 1400 et seq (1975).
- Elliott, S. N. (2007). Selecting and using testing accommodations to facilitate meaningful participation of all students in state and district assessments. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 1-12). Arlington, VA: Council for Exceptional Children.
- 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*(2), 136-150.
- Fuchs, L. S., & Fuchs, D. (1999). Fair and unfair testing accommodations. *School Administrator, 56*, 24-29.
- Fuchs, L. S., Fuchs, D., & Capizzi, A. M. (2005). Identifying appropriate test accommodations for students with learning disabilities. *Focus on Exceptional Children, 37*(6), 1-8.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about testing accommodations. *Exceptional Children, 67*, 67-81.
- Georgia Department of Education. (2001). *Reading Criterion-Referenced Competency Tests Content Descriptions*. Atlanta, GA: Author.

Geisigner, K. F. (2007). Preface. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. ix-xii). Arlington, VA: Council for Exceptional Children.

Geisigner, K. (1994). Psychometric issues in testing students with disabilities. *Applied Measurement in Education* 7(2), 121-141.

Gilberston Schulte, A. A., Elliott, S. N., & Kratochwill, T. R. (2000). Educators' perceptions and documentation of testing accommodations for students with disabilities. *Special Services in the Schools*, 16, 35-56.

Green, P. C., & Sireci, S. G. (1999). Legal and psychometric policy considerations in the testing of students with disabilities. *Journal of Special Education Leadership*, 12(2), 21-29.

Hollenbeck, K. (2005). Validity issues and decisions about testing accommodations. *Assessment for Effective Interventions*, 31, 1, 7-18.

Hollenbeck, K. (2002). Determining when test alterations are valid accommodations or modifications for large-scale assessments. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, Technical Adequacy, and Implementation* (pp. 395-425). Mahwah, NJ: Lawrence Erlbaum Associates.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Bray, G. B., Naylor, R. J., Lewis, J. C., Ordman, V. L., Qualls, A. L. (2003). *The Iowa Tests: Interpretive Guide for School Administrators*. Itasca, IL: Riverside Publishing.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Ordman, V. L., Naylor, R.

J., Bray, G. B., Lewis, J. C., Qualls, A. L., Mengeling, M. A., and Shannon, G. P. (2003). *The Iowa Tests: Guide to Research and Development*. Itasca, IL: Riverside Publishing.

Individuals with Disabilities Education Act (IDEA), as amended in 2004, PL 108-446, 20 USC 1400 et seq.

Individuals with Disabilities Education Act of 1997, Pub. L. No. 105-17, 111 Stat. 37 (1997).

Kettler, R. J., Niebling, B. C., Mroch, A. A., Feldman, E. S., Newell, M. L., Elliott, S.N., Kratochwill, T. R., & Bolt, D. M. (2005). Effects of testing accommodations on math and reading scores: An experimental analysis of the performance of students with and without disabilities. *Assessment for Effective Interventions*, 31, 1, 37-48.

Laitusis, C. C. (2010). Examining the Impact of Audio Presentation on Tests of Reading Comprehension. *Applied Measurement in Education*, 23, 153-167.

Laitusis, C. C. (2008). State reading assessments and inclusion of students with dyslexia. *Perspectives on Language and Literacy*, 34(4), 31-33.

Laitusis, C. C. (2007). Research designs and analysis for studying accommodations on assessments. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 67-80). Arlington, VA: Council for Exceptional Children.

Malouf, D. B. (2005). Accommodation research to guide practice: Comments on what we know and what we need to know. *Assessment for Effective Interventions*, 31, 1, 79-83.

McKevitt, B. C., & Elliott, S. N. (2003). Effects and perceived consequences of using

read-aloud and teacher-recommended accommodations on a reading achievement test. *School Psychology Review*, 32, 583-600.

Meloy, L. L., Deville, C., & Frisbie, D. A. (2002). The effects of a read-aloud accommodation on test scores of students with and without learning disability in reading. *Remedial and Special Education*, 23(4), 248-255.

Messick, S. (1989). Validity. In R. Linn (Ed.) *Educational measurement* (3rd ed., pp. 13-103). Washington, DC: American Council on Education.

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

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Phillips, S. E. (2002). Legal issues affecting special populations in large-scale testing programs. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, Technical Adequacy, and Implementation* (pp. 395-425). Mahwah, NJ: Lawrence Erlbaum Associates.

Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education*, 7, 2, 93-120.

Pitoniak, M. J., & Royer, J. M. (2001). Testing accommodations for examinees with disabilities: A review of psychometric, legal and social policy issues. *Review of Educational Research*, 71(1), 53-104.

Pullin, D. (2007). Accommodations in testing: Law, policy, and practice. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 37-50). Arlington, VA: Council for Exceptional Children.

Randall, J. & Engelhard, G. (2010). Performance of students with and without

disabilities under modified conditions. *The Journal of Special Education*, 44(2), 79-93.

Roeber, E. D. (2002). Appropriate inclusion of students with disabilities in state accountability systems. Denver, CO: Education Commission of the States. Retrieved June 30, 2005, from World Wide Web: <http://www.ecs.org/clearinghouse/40/11/4011.htm>.

Rooney, P. (2011, August). Understanding the population. Presentation at the Race to the Top Assessment Program Technical Assistance Public Meeting for Creating Valid, Reliable, and Fair Assessments for Students with Disabilities and English Learners, Washington, D.C.

Section 504 of the Rehabilitation Act of 1973, 29 U.S.C. § 701 *et seq.*

Sireci, S. G., & Pitoniak, M. J. (2007). Assessment accommodations: What have we learned from research? In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 53-67). Arlington, VA: Council for Exceptional Children.

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75(4), 457-490

Tindal, G., Heath, B., Hollenbeck, K., Almond, P., & Harniss, M. (1998). Accommodating students with disabilities on large-scale tests: An experimental study. *Exceptional Children*, 64, 439-450.

Thurlow, M. (2011, August). What is the current state of research on accommodations

and what future research is necessary? Presentation at the Race to the Top Assessment Program Technical Assistance Public Meeting for Creating Valid, Reliable, and Fair Assessments for Students with Disabilities and English Learners , Washington, D.C.

Thurlow, M. (2010). Steps toward creating fully accessible reading assessments. *Applied Measurement in Education*, 23, 121-131.

Thurlow, M. L. (2007). State policies and accommodations: Issues and implications. In C. C. Laitusis & L. L. Cook (Eds.). *Large-Scale Assessment and Accommodations: What Works?* (pp. 13-22). Arlington, VA: Council for Exceptional Children.

US Department of Education, Office of Elementary and Secondary Education. (2007b). *Modified achievement standards: Non-regulatory guidance*. Washington, DC: Author.

Zenisky, A.L. & Sireci, S.G. (2007). *A Summary of the Research for the Effects of Test Accommodations: 2005 through 2006*. Center for Educational Assessment, University of Massachusetts-Amherst.

APPENDIXES

Appendix A



(404) 656-2668 FAX (404) 656-5976

January 31, 2006

MEMORANDUM

TO: Selected Superintendents
Assessment Directors
Special Education Directors

FROM: Ida Love, Ph.D.
Deputy State Superintendent, Curriculum and Instruction

SUBJECT: Georgia Modification Research Study

As many of you are aware, the US Department of Education (US DOE) has recently announced additional flexibility for students with disabilities. This new flexibility allows states to utilize modified assessments for approximately two percent of the population in the tested grades. States may design modified assessments for students with disabilities who are ineligible for alternate assessments based on alternate achievement standards (i.e., the Georgia Alternate Assessment) but whose disabilities preclude them from achieving grade level achievement in the same timeframe as their peers.

The Georgia Department of Education (GaDOE) has scheduled a small research study in effort to better understand how test modifications affect the reliability and validity of the Criterion-Referenced Competency Tests (CRCT) and the Georgia High School Graduation Tests (GHSGT). Test modifications differ from traditional test accommodations in that they typically alter what the test is measuring. In the majority of cases, modifications involve the addition of content or resources (such as a calculator on a test that is measuring computational skill).

You are receiving this memorandum because a sample of your system's schools has been selected to participate in this important study. Please know that your participation is vital to the success of this study and ultimately to providing more appropriate assessments for our students with disabilities.

Participating in this study will involve administering the CRCT or GHSGT to both special and regular education students in grades 4, 7, and 12. Only the Reading and

Mathematics portions of the CRCT and the English/Language Arts and Mathematics portions of the GHSGT will be administered for this study. It is anticipated that five hours will be the maximum time of student testing for this study.

A complete list of the selected schools is provided as an attachment. The list also provides the number of special and regular education students to be tested. Generally speaking, all students with disabilities in the targeted grades (4, 7, and 12) should participate. The same number of regular education students in these grades should be randomly selected by the school to participate. All students within a school will take the test under the same condition (either a traditional administration or a modified condition). Conditions have been assigned to each school.

The schedule for elementary and middle schools participating in this important study is as follows:

- Test materials (test booklets, answer sheets, and administration directions) will be shipped to arrive in participating school districts by *February 27*.
- Tests should be administered between *March 6 and March 17*.
- All test materials must be returned to the Georgia Center for Assessment at the University of Georgia by *March 22*. Return shipping labels will be provided.

The schedule for high schools participating in this important study is as follows:

- Test materials (test booklets, answer sheets, and administration directions) will be shipped to arrive in participating school districts by *March 29*.
- Tests should be administered between *April 10 – April 21*.
- All test materials must be returned to the Georgia Center for Assessment at the University of Georgia by *April 26*. Return shipping labels will be provided.

For districts participating in this study, the Testing Division has scheduled a pre-administration workshop to be held via WebEx to explain the logistics:

- February 21, 2006, from 1:00 – 3:30; login information will be sent to System Test Coordinators under separate cover.

If you need additional information regarding the research study, please contact Melissa Fincher by phone at 404-651-9405 or by email at mfincher@doe.k12.ga.us.

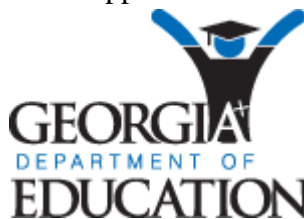
I sincerely thank you for your involvement in this study and appreciate your willingness to help Georgia build an assessment program that allows all students to demonstrate their knowledge and skill.

IL/mlf

Attachment

cc: Kathy Cox, State Superintendent of Schools

Appendix B



**MODIFICATION RESEARCH STUDY
SPRING 2006
SUPPLEMENTAL INFORMATION FOR PARTICIPATING
ELEMENTARY AND MIDDLE SCHOOLS**

STUDY RATIONALE

Recently the US Department of Education (US DOE) announced new flexibility for students with disabilities. Recognizing that not all students are appropriately assessed through a state's regular assessment program (for Georgia, the Criterion-Referenced Competency Tests) or the alternate (the Georgia Alternate Assessment), states may design modified assessments for 2% of the population. Although additional guidance will be forthcoming, the US DOE has described the group of students eligible for modified assessments as those students whose disabilities preclude them from attaining grade-level achievement in the same timeframe as their peers. Instruction and assessment must be on grade-level. States may design modified assessments and develop detailed participation criteria.

The primary purpose of this study is to investigate the effects of selected modifications on the reliability and validity of the Criterion-Referenced Competency Tests (CRCT). Modifications differ from traditional test administration accommodations in that they typically result in a change in the construct measured by the assessment. For example, reading the passages on the CRCT changes the type of comprehension measured. This is a modification rather than an accommodation. Test administration accommodations are changes in the manner in which a test is presented (Braille or large print), how it is administered (additional time), or how a student responds (marking answers in the test book). Accommodations do not change the construct measured by the test. For example, allowing a student additional time on a reading test does not change the type of comprehension measured.

While the Georgia Department of Education (GaDOE) makes every attempt to build assessments that are accessible to all students, some students still struggle with demonstrating what they have learned. Based on requests from school districts, this study is designed to help illuminate which modifications are most appropriate and/or effective.

The study will focus on reading and mathematics tests only. Therefore, students participating in the study should be administered only the Reading and Mathematics CRCT. **Do not administer the English/Language Arts, Science, or Social Studies tests.**

This study involves a sample of schools serving grades 4 and 7, randomly selected to reflect statewide demographics. For security reasons, students will be administered an off-grade level test. This means that students in grade 4 will take the grade 3 CRCT and students in grade 7 will take the grade 6 CRCT. It is important to note that off grade-level testing is not permitted by the US DOE as an acceptable form of modified assessments.

This study will utilize the test forms administered in Spring 2005. The forms are still secure and all security procedures and protocols should apply. Teachers should use the *Examiner's Manual* to administer the tests and follow all procedures in the manual.

MODIFICATION STUDY LOGISTICS

The study will involve both special education students and regular education students. The number of special education students to be tested in grades 4 and 7 has been identified by GaDOE and provided to the System Test Coordinator. (Please note that students who participate in the Georgia Alternate Assessment should not participate in this study.) Each school should randomly select an equal number of regular education students to participate.

Each school has been assigned one of three conditions. All students (both regular and special) within the school will test under the same condition. Two of the three conditions involve the use of modifications; the third condition involves a traditional administration. The three conditions are explained below in more detail.

Condition 1 – Resource Guides: Schools assigned condition 1 will be provided with Resource Guides to use during the assessment. Each content area (reading and math) has a single page (front and back) guide that provides key definitions, examples, graphics, etc., students may find helpful. Short sample tests have been provided for schools assigned this condition. Before beginning the assessment, students should be given the opportunity to work through the sample test with the Resource Guide. Teachers should review the sample tests with the students and, if necessary, provide pointers on how the questions on the sample test relate to information on the guide. Because the idea of a Resource Guide will be new to the majority of students, it is recommended that this activity be done a day or two before the administration.

All students in a school assigned this condition should be provided the Resource Guide. This includes both special education and regular education students. It is very likely that some students will find the guides distracting and/or cumbersome. This should be expected; students should be encouraged to do their best work if they comment about the guide. The guide is provided as a resource; some students may not need or want to utilize the guide. Teachers are asked to provide feedback on the appropriateness of the guides and how their students responded to them. Comments from teachers should be submitted with returned materials.

Special education students may be allowed other additional accommodations required by the student (and in the IEP), provided those accommodations result in a **standard** administration (see pages 31 – 33 of the *Examiner's Manual* for a list of accommodations and an indication of which result in a standard or non-standard administration). *Additional* accommodations that result in a non-standard accommodation should not be provided during this study.

Finally, it is important to note the use of these guides will not be allowed on the Spring 2006 CRCT. All guides should be collected and returned with test materials at the conclusion of the study.

Condition 2 – Read Aloud/Calculator: Schools assigned Condition 2 will provide two modifications – one for the reading test and one for the math test. During the Reading CRCT, the teacher should read the entire test to students, including reading passages and questions. All students within the group should be provided the same test form; materials were packed accordingly, but schools should verify that each test is the same form prior to beginning the administration. The teacher should read the test verbatim being careful to pace the test at a natural rate. The test should be read to both special and regular education students.

Basic function calculators have been provided for the math test. Each student should be given a calculator and allowed to practice use of the calculator, if they are unfamiliar, before beginning the test. Teachers may create problems for students to practice as no sample test is provided for this condition. Teachers should not use items from the Spring 2005 CRCT to practice, but rather create their own items or use items from the Online Assessment System. The math test items should not be read to students; use of calculators is the modification allowed for this condition. Please note that schools may keep the calculators at the conclusion of the study; it is not necessary to return calculators.

Special education students may be allowed other accommodations provided those accommodations result in a **standard** administration (see pages 31 – 33 of the *Examiner's Manual* for a list of accommodations and an indication of which result in a standard or non-standard administration). *Additional* accommodations that result in a non-standard accommodation should not be provided during this study.

Condition 3 – Traditional Administration: Schools assigned Condition 3 will administer the tests traditionally, as if it were a normal administration. Essentially, schools assigned Condition 3 will serve as a comparison group. For this reason, it is very important that the tests be administered appropriately. Regular education students should test without modifications or accommodations. Special education students who require accommodations should be given those accommodations listed in their IEP, provided the accommodations do not result in a non-standard administration. **For the purposes of this research study, only accommodations resulting in a standard administration should be provided** (see pages 31 – 33 of the *Examiner's Manual* for a list of

accommodations and an indication of which result in a standard or non-standard administration). Students should be encouraged to do their best work.

SPECIAL EDUCATION SURVEY

A special survey has been created for special education students participating in this study. This survey was adapted, with permission, from one used by the National Center for Educational Statistics (NCES) for the National Assessment of Educational Progress (NAEP). The educator most knowledgeable about the student should complete the survey. **A survey should be completed for each special education student who participates in the study.** This important information will help identify the characteristics of students who benefit from modifications and inform the development of participation criteria for modified assessments.

ADMINISTRATIVE PROCEDURES

Examiners should follow all administrative procedures set forth in the *Examiner's Manual*. All student demographics should be completed, including State Required Codes. Once testing is complete, all documents for a school should be placed under a header/transmittal form provided by the Georgia Center for Assessment (formerly the Test Scoring and Reporting Services at the University of Georgia).

All materials are secure and must be accounted for daily. Material should be returned to the School Test Coordinator at the conclusion of testing each day. All materials, including Resource Guides, must be accounted for and returned to the System Test Coordinator at the conclusion of the study. Remember to include the Special Education Surveys and any teacher comments about how students responded to the condition assigned to the school. System Test Coordinators will verify materials and return to the Georgia Center for Assessment using the return labels provided.

Your participation in this study is vital to the development of better assessment methodologies for many students across our state. We sincerely appreciate your assistance. Results of the study will be shared with System Test Coordinators in Fall 2006.

Appendix C



**Georgia Department of Education
Special Assessment Study / February – March 2006
Special Education Questionnaire**

(to be completed by the educator most knowledgeable about the student)

Directions Thank you for participating in this special study. The purpose of this study is to help determine the effects certain modifications, as opposed to accommodations, have on state-mandated assessments. Both special education and regular education students will participate in this research study. Systems and schools were randomly selected and assigned a condition of no modification or one of two modifications.

This questionnaire should be completed for every *special education* student participating in the study. The educator most knowledgeable about the student and the student's instructional program should complete this questionnaire. Answers will help determine which students most benefit from modifications.

Thank you for your time and effort.

Student Name: _____

Student FTE Number: _____

Gender: ___ Female ___ Male

State System/School Code: _____ - _____
 System School

For the items that follow, indicate your response by placing a check mark (T) or an X (X) in the blank in front of your choice. Several items ask you to mark all options that apply.

What is your relationship to this student?

- ___ Classroom (General Education) Teacher
- ___ Special Education Teacher
- ___ Related Service Provider (e.g., Speech Language Pathologist, Occupational Therapist, Physical Therapist)
- ___ Guidance/School Counselor
- ___ Principal/Assistant Principal
- ___ Other (specify) _____

- 1) What is this student's identified disability(ies)? (Mark all that apply.)
- Visual impairment
 - Deaf/Hard of Hearing
 - Deaf/Blind
 - Specific Learning Disability
 - Mild Intellectual Disabilities
 - Traumatic Brain Injury
 - Moderate/Severe/Profound Intellectual Disabilities
 - Autism
 - Orthopedic impairment
 - Speech-language impairment
 - Emotional and Behavioral Disorder
 - Other Health Impairments
 - Significant Developmental Delay (K-3 only)
- 2) In your judgment, what is the degree of this student's disability(ies)?
- Profound/Severe
 - Moderate
 - Mild
- 3) At a minimum, do this student's long-term mathematics objectives include the ability to perform basic mathematics calculations without the use of a calculator?
- Yes
 - No
- 4) At a minimum, do this student's long-term reading objectives include the ability to decode simple printed material?
- Yes
 - No
- 5) What proportion of this student's academic class time (in subjects such as mathematics, reading/language arts, science, and social studies) is spent in a mainstream/general education classroom?
- None
 - Half or less
 - More than half, but not all
 - All
- 6) In which area(s) is this student currently receiving special education services? (Mark all that apply.)
- This student does not currently receive special education services.
 - Language development
 - Reading
 - Mathematics
 - Science
 - Social Studies

- Speech (e.g. articulation, voice, speech flow)
- Self-control and/or deportment
- Personal care and/or basic life skills
- Vocational education
- Other (specify) _____

Questions 7-13 ask about this student's instruction, assessment, and accommodations in **reading**.

Instruction and Assessment

- 7) What grade level of instruction is this student currently receiving in **reading**?
- This student is currently not receiving instruction in this subject.
 - At or above grade level
 - One year below grade level
 - Two or more years below grade level
- 8) Is this student participating in the same curriculum content as nondisabled students in **reading**?
- This student is currently not receiving instruction in this subject.
 - Same curriculum content.
 - Different curriculum content
- 9) According to the student's IEP, how does this student participate in the state academic assessment in **reading**?
- Regular assessment without accommodations
 - Regular assessment using accommodations that result in a standard administration
 - Regular assessment using accommodations that result in a nonstandard administration
 - Alternate assessment for students who are significantly cognitively disabled
 - Other (specify) _____

Accommodations

- 10) Presentation Accommodations for **reading** (Mark all that apply.)
- No presentation accommodations
 - Directions read aloud to student
 - Directions signed
 - Directions repeated or paraphrased/explained
 - Assistance with interpretation of directions given
 - Passages and test questions read aloud or signed
 - Braille edition of test
 - Large-print edition of test
 - Magnifying equipment provided
 - Test administered by person familiar to the student

- Color overlays or templates
- Optimal time of day for testing
- Other (specify) _____

11) Response Accommodations for **reading** (Mark all that apply.)

- No response accommodations
- Responds in Braille
- Responds in sign language
- Points to answers
- Responds orally
- Uses word processor (with grammar and spell check devices disabled)
- Adapted writing tool (e.g., pencil grips, large pencils)
- Writes directly in test booklet
- Calculator
- Other (specify) _____

12) Setting Accommodations for **reading** (Mark all that apply.)

- No setting accommodations
- Tested in small group
- Tested individually
- Preferential seating
- Special or adapted lighting
- Adapted furniture (e.g., slant board)
- Individual study carrel
- Other (specify) _____

13) Scheduling Accommodations for Reading (Mark all that apply.)

- No scheduling accommodations
- Extended time
- Frequent monitored breaks during test
- Optimal time of day for testing
- Other (specify) _____

Questions 14-20 ask about this student's instruction, assessment, and accommodations in **mathematics**.

Instruction and Assessment

14) What grade level of instruction is this student currently receiving in **math**?

- This student is currently not receiving instruction in this subject.
- At or above grade level
- One year below grade level
- Two or more years below grade level

15) Is this student participating in the same curriculum content as nondisabled students in **math**?

- This student is currently not receiving instruction in this subject.
 Same curriculum content.
 Different curriculum content

16) According to the student's IEP, how does this student participate in the state academic assessment in **math**?

- Regular assessment without accommodations
 Regular assessment using accommodations that result in a standard administration
 Regular assessment using accommodations that result in a nonstandard administration
 Alternate assessment for students who are significantly cognitively disabled
 Other (specify) _____

Accommodations

17) Presentation Accommodations for **math** (Mark all that apply.)

- No presentation accommodations
 Directions read aloud to student
 Directions signed
 Directions repeated or paraphrased/explained
 Assistance with interpretation of directions given
 Passages and test questions read aloud or signed
 Braille edition of test
 Large-print edition of test
 Magnifying equipment provided
 Test administered by person familiar to the student
 Color overlays or templates
 Optimal time of day for testing
 Other (specify) _____

18) Response Accommodations for **math** (Mark all that apply.)

- No response accommodations
 Responds in Braille
 Responds in sign language
 Points to answers
 Responds orally
 Uses word processor (with grammar and spell check devices disabled)
 Adapted writing tool (e.g., pencil grips, large pencils)
 Writes directly in test booklet
 Calculator
 Other (specify) _____

19) Setting Accommodations for **math** (Mark all that apply.)

- No setting accommodations
- Tested in small group
- Tested individually
- Preferential seating
- Special or adapted lighting
- Adapted furniture (e.g., slant board)
- Individual study carrel
- Other (specify) _____

20) Timing Accommodations for **math** (Mark all that apply.)

- No scheduling accommodations
- Extended time
- Frequent monitored breaks during test
- Optimal time of day for testing
- Other (specify)