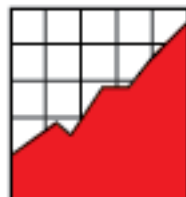


## **A Summary of Research on the Effects of Test Accommodations: 2002 Through 2004**



**NATIONAL  
CENTER ON  
EDUCATIONAL  
OUTCOMES**

***In collaboration with:***

Council of Chief State School Officers (CCSSO)

National Association of State Directors of Special Education (NASDSE)

***Supported by:***

U.S. Office of Special Education Programs

## Technical Report 45

# A Summary of Research on the Effects of Test Accommodations: 2002 Through 2004

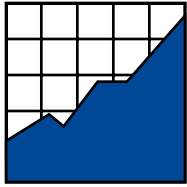
Christopher J. Johnstone • Jason Altman • Martha L. Thurlow •  
Sandra J. Thompson\*

**September 2006**

All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

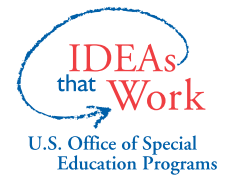
Johnstone, C. J., Altman, J., Thurlow, M. L., & Thompson, S. J. (2006). *A summary of research on the effects of test accommodations: 2002 through 2004* (Technical Report 45). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

\* Dr. Thompson was a Research Associate at NCEO when working on this report. She passed away December 2005 after a career of improving outcomes for students with disabilities.



**N A T I O N A L  
C E N T E R O N  
E D U C A T I O N A L  
O U T C O M E S**

The Center is supported through a Cooperative Agreement (#H326G050007) with the Research to Practice Division, Office of Special Education Programs, U.S. Department of Education. This report was completed under the Center's previous Cooperative Agreement (#H326G000001) with the Office of Special Education Programs. The Center is affiliated with the Institute on Community Integration at the College of Education and Human Development, University of Minnesota. Opinions expressed herein do not necessarily reflect those of the U.S. Department of Education or Offices within it.



**NCEO Core Staff**

Deb A. Albus	Michael L. Moore
Manuel T. Barrera	Rachel F. Quenemoen
Christopher J. Johnstone	Dorene L. Scott
Jane L. Krentz	Karen Evans Stout
Kristi K. Liu	Martha L. Thurlow, Director
Ross E. Moen	

National Center on Educational Outcomes  
University of Minnesota • 350 Elliott Hall  
75 East River Road • Minneapolis, MN 55455  
Phone 612/626-1530 • Fax 612/624-0879  
<http://www.nceo.info>

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.

This document is available in alternative formats upon request.

## Executive Summary

---

The No Child Left Behind Act of 2001 (NCLB) requires the reporting of participation in assessments overall and by subgroup, including students with disabilities. As states and school districts strive to meet the goals for adequate yearly progress required by NCLB, the use of individual accommodations continues to be scrutinized for effectiveness, threats to test validity, and score comparability. This report summarizes 49 empirical research studies completed on test accommodations between 2002 and 2004, and provides direction in the design of critically needed future research on accommodations.

NCEO found that studies during this three-year period had the following characteristics:

**Purpose.** The primary purpose of the 2002-2004 accommodations research was to determine the effects of accommodations use on the large-scale test scores of students with disabilities.

**Types of assessment, content areas, and accommodations.** The majority of the studies tested students using norm-referenced or criterion-referenced tests, on math or reading/language arts.

**Participants.** Equal numbers of research studies involved between 1-100 participants, 100-1,000 participants, and more than 1,000 participants of multiple age categories. Participants were varying percentages of students without disabilities and students with disabilities. Students with learning disabilities were studied most frequently among students who receive special education services.

**Findings.** Findings shared no common theme, with various accommodations shown to have both a positive and non-positive effect on scores. Individual accommodations showed either differential item functioning or no differential item functioning depending on the study. The lack of consistent findings points to a need for further research.

**Limitations.** Most often, authors noted that studies were too narrow in scope, involved a small sample size, or provided confounding factors. These limitations and other considerations led researchers to recommend investigating the characteristics of accommodations in further detail.

Important overall observations from the NCEO analysis include a need in future research for a clear definition of the constructs tested, a reduction in confounding factors, increased study of institutional factors affecting accommodations judgment, and exploration of the desirability and perceived usefulness of accommodations by students themselves. Future research should focus on improvement in these areas but also on the positive effects of field-testing potential items in accommodated formats in addition to standard formats.

## Table of Contents

---

Introduction.....	1
Methods.....	2
Results.....	3
Purpose of Accommodations Research.....	3
Types of Assessment .....	5
Content Areas Assessed .....	6
Type of Accommodation.....	6
Research Participants .....	7
Research Results .....	9
Limitations .....	12
Recommendations for Future Research.....	13
Discussion and Implications for Future Research .....	14
References.....	17
Appendix A—Summary of Research Purpose.....	23
Appendix B—Summary of Type of Assessment .....	31
Appendix C—Subject Area Studied (by Author).....	47
Appendix D—Type of Accommodation Studied (by Author) .....	49
Appendix E—Summary of Participants.....	53
Appendix F—Summary of Research Results .....	57
Appendix G—Summary of Limitations Cited by Researchers.....	63
Appendix H—Summary of Suggestions for Future Research (as recommended by authors) .....	67

## Introduction

---

Over the past decade, students with disabilities have increasingly participated and performed at proficient levels on general education assessments. Participation and proficient performance of all students are required by the No Child Left Behind Act of 2001. As participation rates increase, so does the use of testing accommodations. Increased use of accommodations should reflect an attempt to ensure that the scores received by students with disabilities are valid measures of achievement. It is also possible that increased use of accommodations is simply a reflection of concern about including students in assessments and a belief that these students need additional aids to help them perform better. Because of this, states are clarifying appropriate accommodation use in state policy, with the goal of encouraging Individualized Education Program (IEP) teams to select accommodations that remove specific disability barriers, but do not give students with disabilities an unfair advantage over their peers. States have begun to monitor accommodations use, and this will help them to better track the effects of accommodations. In 2005, 20 states maintained a database of accommodations actually used during testing within the state, and 26 states documented the specific accommodations used by students on test day (Thompson, Johnstone, Thurlow, & Altman, 2005).

State policymakers and practitioners define accommodations in a variety of ways. For the purposes of this report, we draw from accommodations research to help shape our definitions and outlook on testing accommodations for students with disabilities. For example, Thurlow and Bolt (2001) defined testing accommodations as:

changes in assessment materials or procedures that address aspects of students' disabilities that may interfere with the demonstration of their knowledge and skills on standardized tests. Accommodations attempt to eliminate barriers to meaningful testing, thereby allowing for the participation of students with disabilities in state and district assessments. (p. 3)

Sireci, Li, and Scarpati (2005) explained the validity of accommodations through an “interaction hypothesis” or the theoretical assumption that test accommodations will lead to improved test scores for students who need accommodations, but not for students who do not need accommodations (i.e., students with disabilities receive a boost in scores as a result of accommodations whereas students without disabilities do not receive a boost or receive a less pronounced boost in scores).

The accommodations allowed in state assessment policies vary from state to state. The most common accommodations found in previous syntheses of research were read aloud accommodations (sometimes referred to as oral administration), computer administration of tests, extended time and tests across multiple days, calculator use, and use of a scribe (Thompson, Blount, &

Thurlow, 2002). Most students use a combination of accommodations (Bielinski, Ysseldyke, Bolt, Friedebach, & Friedebach, 2001), as dictated by their IEPs.

Research on accommodations has yielded mixed results in terms of validity and efficacy of providing them to students with disabilities. Over the past several years, accommodations research has yet to produce definitive answers for policymakers and practitioners (Sireci, Li, & Scarpati, 2003). Nevertheless, research reports on accommodations continue to be found in professional journals, indicating that there is still a need to investigate the complex issues surrounding accommodations. This report examines empirical research published between 2002 and 2004. We searched peer-reviewed articles, technical reports, and dissertations in order to provide the readers with up-to-date information on accommodations.

In this report, research is summarized according to several components, including research purpose, type of assessment, content area assessed, type of accommodation, number of participants, percent of sample consisting of students with disabilities, participant grade level, type of disability, research results, research limitations, and recommendations for further research. During our review process, we found 49 published studies on accommodations between 2002-2004. This number reflects a high number of studies conducted in the new millennium, and is slightly more than the number of studies published between 1999 and 2001 (Thompson, Blount, & Thurlow, 2002) (see Table 1). The publications reviewed for this report are found in Appendix A.

**Table 1. Number of Accommodations Studies by Years**

<b>Years</b>	<b>Number of Studies</b>
1990 through 1992	11
1993 through 1995	18
1996 through 1998	29
1999 through 2001	46
2002 through 2004	49

## **Methods**

---

NCEO used a four-stage process to find publications related to accommodations from 2002 through 2004. First, we conducted a search of electronic databases including ERIC, PsychInfo, Educational Abstracts, and Digital Dissertations using the keywords “accommodation,” “test adaptation,” “test changes,” “test modifications,” “test accommodations,” “state testing accommodations,” “standards-based testing accommodations,” and “large-scale testing accommodations.”

A second electronic search consisted of organizational Web sites, including Behavior Research and Training (<http://brt.uoregon.edu/>), the National Center for Research on Evaluation, Standards, and Student Testing (<http://www.cse.ucla.edu/>), the Center for the Study of Assessment Validity and Evaluation (<http://www.c-save.umd.edu/index.html>), and the Wisconsin Center for Educational Research (<http://www.wcer.wisc.edu/tesacc/>). In addition, an archival search of Educational Policy Analysis Archives was undertaken.

In addition to the electronic searches, NCEO staff performed two hand searches. First, references from all selected materials dated 2002 through 2004 were examined in an effort to find further source material. Second, 2002 through 2004 issues of major measurement and special education journals were hand searched in the University of Minnesota library. These journals included *Applied Measurement in Education*, *British Journal of Special Education*, *Diagnostique*, *Educational Assessment*, *Educational and Psychological Measurement*, *Educational Measurement: Issues and Practice*, *Educational Psychologist*, *Educational Psychology*, *Exceptional Children*, *Journal of Educational Measurement*, *Journal of Learning Disabilities*, *Journal of School Psychology*, *Journal of Special Education*, *Remedial and Special Education*. Last, the schedules of the annual conferences of major organizations (such as the American Educational Research Association, the Council of Chief State School Officers, Council for Exceptional Children, and the National Council on Measurement in Education) were scanned for presentations on accommodations.

NCEO research staff searched all identified sources over an 18-month period beginning in fall, 2004 and concluding in spring, 2005. All of the studies cited were either empirical research or meta-analyses, each with succinct research findings that added to the field's knowledge about the effects of accommodations. The References section of this report includes all journal article, research report, conference presentation, and dissertation references.

## **Results**

---

### **Purpose of Accommodations Research**

Two primary purposes appeared most often in accommodations research from 2002 through 2004 (see Table 2). First, the majority of studies examined the effect of the use of accommodations on scores; 23 studies sought to determine the effect of the use of accommodations on test scores with students with disabilities, and 13 studies investigated the effects of accommodations on test score validity. Second, a set of studies seemed to reflect the purpose of looking at accommodations institutional factors (such as teacher knowledge, effects of policy, and IEP team decision making; nine publications fit this purpose. In addition to these, two publications examined patterns of errors across items or tests and two meta-analyses synthesized accommodations studies. Details of the studies according to purpose are provided in Appendix A.



**Table 2. Research Purposes**

<b>Research Purpose</b>	<b>Number of Studies</b>
Determine the effect of the use of accommodations on test scores of students with disabilities	23
Investigate the effects of accommodations on test score validity	13
Study institutional factors, teacher judgment, or student desirability of accommodation use	9
Examine patterns of errors across items or tests	2
Meta-analysis	2

Authors employed a variety of methods in their research (see Table 3). The most common methods were experimental and quasi-experimental research, in which research participants took tests under different conditions, and reviews of extant data, in which researchers reviewed data from assessments that students took with or without accommodations that were not specifically designed for experimental and control conditions. Twenty-one publications used experimental design in their methodology and 17 studies reviewed existing data from state large-scale assessments or local assessments. A variety of descriptive and comparative statistics were employed to examine extant data. In addition to these methods, seven additional studies used survey or interview methods to better understand stakeholder understanding of opinions on accommodations. Two studies were meta-analyses, one study evaluated a product and one study described interventions for IEP teams.

**Table 3. Research Methods**

<b>Method</b>	<b>Number of Studies</b>
Experimental or Quasi-experimental	21
Review of extant data	17
Survey/Interview	7
Meta-analysis	2
IEP intervention	1
Product evaluation	1

Note: Studies are described by their primary methodology.

The experimental and extant data analysis methods combined had two main goals: understanding the effect of accommodations on test scores and understanding the effects of accommodations on the psychometric qualities of items. Among the 23 studies that examined the effects of accommodations on test scores (see Table 2), researchers found that computerized administration (Pomplun, Frey, & Becker, 2002), read-aloud accommodation (Helwig, Rozek-Tedesco, & Tindal, 2002; Meloy, Deville, & Frisbie, 2002), video administration (Burch, 2002; Tindal, 2002), extended time (Bridgeman, Cline, & Hessinger, 2004), and assistive technology (Landa, Russell, Gourgey, Erin, & Cowan, 2003; MacArthur & Cavalier, 2004) all had a positive effect on the test scores of at least some of the students with disabilities included in the research

samples. Conversely, Burch (2002) and Barton (2002) found that some students with learning disabilities did not benefit from computer or video accommodations. Likewise, Schuneman, Camara, Cascallar, Wendler, and Lawrence (2002) found that calculator usage did not have an effect on student scores. According to Elliott and Marquart (2003, 2004), extended time also did not yield improved scores for students with disabilities.

In terms of the psychometric properties of accommodations, researchers obtained mixed results in terms of score comparability for items. Barton (2002), Barton and Huynh (2003), Calahan, Mandinach, and Camara (2002), Huynh, Meyer, and Gallant-Taylor (2002), and Kobrin and Young (2003) found no change in item comparability when various accommodations (including read aloud, extended time, and computerized administration) were employed. Bolt and Bielinski (2002), Choi and Tinker (2002), and Thornton, Reese, Pashley, and Dalessandro (2002), however, all found that tests administered orally, with extended time, or via computer changed item difficulty or constructs.

Finally, findings related to the institutional issues around accommodations were also mixed. Differing foci for studies yielded different results. Six studies of teachers, administrators, and students yielded contrasting results on the relative knowledge of school personnel about accommodations. Cisar (2004) found that special education teachers were more knowledgeable, while Gagnon and McLaughlin (2004) found that teachers and administrators scored similarly on knowledge measures. Woods (2004) discovered that students do not often predict their need for accommodations well.

## Types of Assessment

Researchers who studied the effects of accommodations used two main types of assessment to determine effects and error in the use of accommodations in 2002-2004. This information is shown in Table 4. One common approach to testing the effects of accommodations was for researchers to use norm-referenced tests. Education professionals typically use norm-referenced tests for national comparison, diagnostic decisions in schools, and for college entrance decisions. Researchers examined the effects of accommodations on the following norm-referenced tests: California Achievement Tests (CAT), Graduate Record Exam (GRE), Law School Admission Test (LSAT), Nelson-Denny Reading Test, Scholastic Aptitude Test (SAT), Terra Nova, and the General Certificate of Secondary Education Examination-United Kingdom (GCSE-UK).

In addition, researchers employed a number of statewide criterion-referenced tests (including tests from Maryland, Missouri, Oregon, and South Carolina). One researcher gathered descriptive data from a test that was still in the prototype stage. Appendix B provides details of the assessments used in the research during the years 2002-2004.

**Table 4. Type of Assessment**

Type of Assessment	Number of Studies
Norm-referenced and Other Standardized Tests	18
State Criterion-referenced Tests or Performance Assessments	18
School or District-designed Tests	0
Other	8
Survey	3
N/A, Meta-Analyses	2

### Content Areas Assessed

Authors of studies conducted research in five major academic content areas: reading (n=23), mathematics (n=21), science (n=3), writing (n=3), and social studies (n=1) (see Table 5). There were also seven studies that examined test accommodations, but not in a specific content area. The “no specific content area” studies included examinations of test accommodations on general academic assessments, but did not include surveys, or meta-analyses (n=5). Appendix C gives additional details on the content area examined in research studies (11 studies included two or more content areas).

**Table 5. Content Areas Assessed\***

Content Areas Assessed	Number of Studies
Mathematics	21
Reading/Language Arts	23
Science	3
Writing	3
Social Studies	1
No Specific Content Area	7

\*Studies may have reported on multiple content.

### Type of Accommodation

We found 15 types of accommodations in the research literature from 2002 through 2004 (see Table 6). Four groups of accommodations emerged: presentation (n=21), timing/scheduling (n=8), response (n=2), and technological aid (n=2) (see also Appendix D). In addition, 11 of the studies investigated the effects of multiple accommodations.

Presentation accommodations were investigated most frequently in the research from 2002-2004. Among publications about presentation accommodations, studies about oral administration of tests (read aloud accommodations) were most common (n=11). The use of computers as a testing

accommodation was also common (n=5). In addition, researchers investigated video administration of tests (n=2), large print accommodations (n=1), dictionary use (n=1), and braille formats of tests (n=1). Additional studies examined multiple types of accommodations.

Of the eight studies that explored the timing or scheduling of tests, seven studies investigated extended time and one study examined the outcomes of testing over multiple days. Among the two studies about response formats, one study examined student dictated response. In addition, one studied the use of calculators. Two studies investigated technological aids.

Eleven studies examined the effects of multiple accommodations. Studies of multiple accommodations included combinations of accommodations such as read aloud, video presentation, extra time, large print, individual settings or small group settings.

**Table 6. Type of Accommodation**

<b>Type of Accommodation</b>		<b>Number of Studies</b>
Presentation (21):	Oral Administration	11
	Computer Administration	5
	Video	2
	Large Print	1
	Dictionary Use	1
	Braille	1
Timing/Scheduling (8):	Extended Time	7
	Multiple Day	1
Response (2):	Dictated Response	1
	Calculator	1
Technological Aid (2)		2
Multiple Accommodations (11)		11
N / A (Survey or Meta-Analysis) (5)		5

## Research Participants

Studies varied in the number of participants included in the sample. Table 7 shows the number of research participants in studies reflected in intervals of 100 participants. Eight high school and college students participated in the smallest study (Landau et al., 2003), and Hall (2002) used test data from 192,000 students in the largest study. Full details of research participant numbers is provided in Appendix E. Overall, approximately one-third of the studies had 99 or fewer participants (n=18); approximately one-third had between 100-999 participants (n=13), and approximately one-third of studies has more than 1,000 research participants (n=16).

**Table 7. Number of Participants in Studies**

<b>Number of Participants</b>	<b>Number of Studies</b>
1-99	18
100-199	2
200-299	3
300-499	4
500-999	4
More than 1000	16
Not Applicable	2

The percentage of research participants with disabilities differed across studies (see Table 8). In 18 studies, students with disabilities made up a majority of the sample (participants in eight studies were 50-74 percent students with disabilities and the samples in ten studies were 75-100 percent students with disabilities). In 15 studies, students with disabilities comprised less than half of the sample (there were less than 25 percent students with disabilities in 10 studies and between 25 and 49 percent students with disabilities in the sample of five studies). Nine studies did not report the percentage of the sample with disabilities and seven studies (including two meta-analyses) did not use research methods that involved students.

**Table 8. Percent of Sample Consisting of Students with Disabilities**

<b>Percent of Sample Consisting of Students with Disabilities</b>	<b>Number of Studies</b>
1-24%	10
25-49%	5
50-74%	8
75-100%	10
Not Reported	9
No Students with Disabilities Participated in Study	0
Not Applicable	7

The grade level of research participants also varied (see Table 9). For example, six studies targeted students who were in elementary school (grades K-5), six studies examined accommodations with middle school students (grades 6-8), and 11 studies examined accommodations with high school students (grades 9-12). Postsecondary students participated in six studies and 15 studies investigated students across grade levels (from grades K to 12). Five studies employed surveys or were meta-analyses that did not involve actual research participants.

**Table 9. Grade Level of Participants in Studies**

<b>Participant Grade Level</b>	<b>Number of Studies</b>
Elementary (K-5)	6
Middle School (6-8)	6
High School (9-12)	11
Multiple Grade Level Categories (K-Postsecondary)	15
Post Secondary	6
Not Applicable	5

In addition to differences in grade level, students with a variety of disability labels participated in studies (see Table 10). Several studies (n=16) included more than one disability category. Fifteen studies included students with learning disabilities, 12 studies included students with communication disabilities, 10 studies included students with cognitive disabilities, and 10 studies included students with emotional/behavioral disabilities. Students with less common disabilities, including physical impairment, sensory disabilities, autism, attention deficit disorder, health impairments, and multiple disabilities, were each included in at least one study. Twenty-one studies did not report the types of disabilities of participants.

**Table 10. Disability Categories Included in Studies**

<b>Type of Disability</b>	<b>Number of Studies</b>
Learning Disability	15
Cognitive Disability (e.g., mental retardation)	10
Emotional/Behavioral Disability	10
Communication Disability	12
Reading or Math Deficit	4
Other (includes physical and sensory disabilities, autism, attention deficit disorder, health impairments, and multiple disabilities)	16
Not Reported	21
Not Applicable	5

Note: Studies sometimes include students with more than one disability category; all are reflected in this table.

## Research Results

Results from the 49 studies reviewed in this synthesis varied. Researchers found accommodations showed both statistically positive and statistically non-significant effects on scores. Likewise, some accommodations had no effect on item comparability, while other types of accommodations compromised item comparability.

Presented here are results according to the type of accommodation used. These results are shown in Table 11, with detailed results available in Appendix F. Similar to previous reviews of accommodations research (Thompson, Blount, & Thurlow, 2002), accommodations appeared to have mixed effects in studies from 2002 through 2004. Furthermore, even when separated by type of accommodation, studies still demonstrated mixed effects. The results from the 11 studies that investigated multiple accommodations are not synthesized due to lack of study focus comparability.

**Oral Presentation (Read Aloud).** A total of 11 studies on oral administration of assessments (often called “read aloud” accommodations) produced mixed results. For example, Helwig et al., (2002), Huynh, Meyer, and Gallant (2004), Janson (2002), Meloy, Deville, and Frisbie (2002), Tindal (2002), and Weston (2003) all found that read aloud accommodations had a positive effect on scores for students with disabilities. Bolt and Bielinski (2002) and McKeivitt and Elliott (2003), however, found read aloud accommodations had no significant impact on student scores.

In terms of item comparability, Barton (2002) and Barton and Huynh (2003) found that items were comparable, whether presented under standard conditions or with read aloud accommodations. However, Bolt and Bielinski (2002), Meloy et al. (2002), and Weston (2003), found that read aloud accommodations did affect item comparability.

The determination of who should use oral presentation accommodations is also an issue. Woods (2004) found great inaccuracy in self-prediction for the need of the read aloud accommodation. Such disparate findings point to the on-going controversies regarding read-aloud accommodations.

One study examined the impact of a student-reads-aloud (i.e., student reads text but aloud) accommodation on the performance of middle and high school students with and without learning disabilities on a test of reading comprehension. Elbaum, Arguelles, Campbell, and Saleh (2004) discovered that students’ test performance did not differ in the two conditions, and students with learning disabilities did not benefit more from the accommodation than students without learning disabilities.

**Extended Time.** Authors of eight studies examined how the use of extended time affected student achievement levels on tests and the extent to which items under extended time or multiple day administrations of tests compared to those administered under standard conditions. Several studies (Bridgeman et al., 2004; Dempsey, 2004) found that students with disabilities profit from extended time accommodations. In these studies, students with disabilities had higher test scores because of extended time accommodations.

Buehler (2002) and Elliott and Marquart (2004), however, found no significant effect on scores when students were provided extended time. Such disparities in study results again demonstrate the lack of consistency in accommodations research. The comparability of specific items under different administration categories further complicates accommodations issues. Buehler (2002) found varying student results for items that were deemed to be comparable under standard and extended time administrations, but Elliott et al. (2004) and Thornton et al. (2002) found that the items they studied were not comparable under different administrations. A study by Crawford, Helwig, and Tindal (2004) of multi-day testing produced conflicting results.

**Computer Administration.** In total, five studies investigated the use of computer administered tests from 2002-2004. Among these studies, Pomplun, Frey and Becker (2002) found that students had positive test results when administered tests via computer (rather than paper and pencil format). Barton and Huynh (2003), Bridgeman, Lennon and Jackenthal (2003), and Kobrin and Young (2003), however, found that computer administration of tests had no statistical effect, or a statistically negative impact on student scores. Related studies on item comparability between computer and paper/pencil administration yielded similar results, with Choi and Tinker (2002) determining that items were changed as a result of format differences.

**Technological Aid.** In the years 2002 through 2004, three research studies investigated technology-based testing accommodations, and all resulted in positive effects on test scores. Hansen, Lee, and Forer (2002) found that the usability of speech output technology was evaluated positively, and that ‘self-voicing’ testing systems have significant potential and may be capable of replacing human readers in certain testing situations. Landau et al. (2003) found that the Tactile Text Tablet, a hybrid between a braille paper-based test and laptop computer, had positive effects on student achievement. MacArthur and Cavalier (2004) found that the use of speech recognition software was feasible, created impressive dictation results, and improved the quality of student-written essays.

**Calculator Use.** There was only one study between the years of 2002-2004 that considered calculator use. In this study Scheuneman et al. (2002) found that calculator use had no significant effect on scores for students taking the SAT. It is unknown why research in calculator usage has become less common than in past years (there were four research studies on calculator accommodations in the years 1999-2001 but only one accommodations study during the years of 2002-2004).

**Dictionary Use.** One international study of dictionary accommodations took place from 2002 through 2004. Idstein’s (2003) study of Israeli students found that dictionaries were not an effective accommodation and that dictionary use interrupted student thought patterns.



**Table 11. Types of Accommodations in Studies**

Type of Accommodation	Research Results	Number of Studies
Oral Presentation (11)	Positive effect on scores	6
	No Differential Item Functioning	2
	No significant effect on scores	2
	Differential Item Functioning	2
	Self-prediction for need unreliable	1
Extended Time (12)	Positive effect on scores	3
	No significant effect on scores	2
	Differential Item Functioning	2
	No Differential Item Functioning	1
	Scores on accommodated test predictor of grades	1
Computer Administration (5)	No significant effect on scores	3
	Positive effect on scores	1
	Differential Item Functioning	1
Technological Aid (3)	Positive effect on scores	3
Calculator Use (1)	No significant effect on scores	1
Dictionary Use (1)	Negative effect on Scores	1
Multiple Accommodations		11
N/A, Meta-analyses, Survey, Teacher		5

## Limitations

Educational research has inherent limitations that require readers to consider findings carefully. For example, true experimental conditions rarely mimic the true conditions in schools under “live” testing conditions. In addition, sample sizes for studies of students with disabilities are often small because students with disabilities are a minority population in schools (roughly one in 10 students has a disability).

Thirty-six authors (74 percent) of accommodation studies published from 2002 through 2004 noted limitations in their studies. Table 12 provides tabular information on most commonly found limitations, and Appendix G provides brief annotations of studies that reported limitations.

Fifteen authors reported a small or narrow sample size, including Hall (2002) who noted, “The study focuses only on fifth grade students, and the results may not generalize to students with disabilities in other grades or dissimilar disabilities, socio-economic statues, etc.” Thirteen authors warned that confounding factors may have influenced results. Common factors were testing multiple accommodations at once and an inability to randomize the sample. Four authors found a flaw in research design that affected study results. Two authors each listed conflicting

results or nonstandard administration across proctors and schools as a limitation. Eleven authors did not mention a limitation nor did the two meta-analyses.

**Table 12. Limitations of Research**

<b>Research Limitation</b>	<b>Number of Studies</b>
Small Sample Size/Sample Too Narrow in Scope	15
Confounding Factors	13
Flaw in Research Design	4
Conflicting Results	2
Nonstandard Administration Across Proctors and Schools	2
No Limitations Mentioned	12
Not Applicable/Meta-Analyses	2

### Recommendations for Future Research

From 2002 through 2004, 34 research studies included recommendations for further research. Table 13 represents the categories of recommendations listed by researchers, with more detailed explanations available in Appendix H. Calls for further investigation demonstrate the continuing investigatory nature of accommodations research. Although scholars have conducted accommodations research for several decades, there is still a clear need for more examination in various areas. In studies conducted from 2002 through 2004, authors suggested there was need for further understanding of the factors of accommodations that contribute to possible variation in results, further understanding of student factors that contribute to accommodations use and success, improved study design or replication of studies, further research on accommodations policy and overall hypotheses, replication of studies due to small sample sizes, investigation into teacher characteristics that relate to accommodation selection, investigation into the possible uses for accommodations in instruction, and investigation into accommodation use practicality.

**Table 13. Recommendations for Future Research**

<b>Recommendations</b>	<b>Number of Studies</b>
Investigate characteristics of accommodations themselves in further detail	12
Investigate student factors contributing to accommodations use	6
Improved study design or study replication	5
Study policy and accommodations hypotheses	4
Replicate study with larger sample	3
Investigate teacher factors related to accommodations selections	2
Investigate possible instructional uses for accommodations	1
Investigate accommodation practicality	1

## Discussion and Implications for Future Research

Several themes arose from the 49 studies of accommodations published between 2002 and 2004. One theme was that accommodations research is inconclusive. This is similar to past findings from NCEO summaries of research (Thompson, Blount, & Thurlow, 2002). Given that there is not a preponderance of evidence concerning accommodations, nearly two-thirds of the authors (n=31) suggested that future research is needed to solidify understanding of accommodation effects.

Researchers published three more accommodations studies from 2002 through 2004 than from 1999 through 2001. Similar to previous years, the majority of studies in the most recent period focused on test scores of students with disabilities related to accommodations. A significant number of studies also investigated the effects of accommodations on test score validity. Researchers were particularly concerned about accommodations changing the construct of the items assessed. Similar to previous reviews, a smaller number of studies from 2002 through 2004 concentrated on institutional factors related to accommodation use (e.g., teacher judgment, student selection, policy), on patterns of errors across items, or were meta-analyses of previous work.

The assessments that researchers selected for examination were primarily standardized, norm-referenced tests and performance assessments. As would be expected with assessment requirements of the No Child Left Behind Act of 2001, most accommodations research was conducted in the areas of reading/language arts and mathematics. Across subject areas, researchers studied oral administration and timing accommodations most often.

Sample sizes varied, with approximately equal representation of small (n=99 or fewer subjects), medium (n=100-999 subjects), and large (n≥1,000 subjects). Studies with low sample sizes typically targeted students with disabilities in K-12 schools while large studies typically used tests administered to large numbers of students (such as college entrance examinations). Despite the uniform nature of sample sizes, there was less uniformity in terms of the grade-level of participants. Accommodations studies were spread across grade and educational level, but were conducted primarily with secondary and post-secondary education students.

Finally, in terms of demographics, researchers most often studied students with learning disabilities from 2002 through 2004. Eight studies explicitly targeted students with learning disabilities. This pattern mirrors accommodations patterns from 1999-2001 (Thompson, Blount, & Thurlow, 2002) most likely because students with learning disabilities are the largest group of students with disabilities and because this population is frequently assigned accommodations such as oral administration and extended time. The practical value of studying these accommodations with students with learning disabilities is obvious, and was evident in research from 2002 through 2004.

Although this report is not meant to provide a scientific meta-analysis of accommodations research (for meta-analyses see Sireci et al., 2005 and Tindal and Ketterlin-Geller, 2004), general patterns that emerged from a review of accommodations research in the years 2002, 2003, and 2004 indicate possible considerations for future research.

The majority of research concentrated on the effect of accommodations use for students with disabilities and the effects on score validity due to accommodations use. Although there were 36 studies combined that investigated scoring and validity, there was little consensus among researchers. Findings continue to be contradictory. Research indicated that accommodations were either beneficial or not beneficial for students with disabilities. Likewise, researchers did not reach consensus on whether accommodations change the construct of the item assessed. Because findings are relatively disparate, there does appear to be a need for further research.

Research that continues to delineate the “interaction hypothesis” (Sireci et al., 2005) and that reduces construct irrelevant variance for students with disabilities without introducing any new effects for non-disabled students still appears to be necessary. In 2002-2004, 21 studies employed experimental or quasi-experimental methods. Replications of scientific methods to discover the effects of accommodations may help the field to better understand how accommodations effect scoring and validity.

While scientific research holds great importance for scoring and validity issues, the variety of research conducted over the course of 2002-2004 is also important. Studies in 2002, 2003, and 2004 investigated accommodations score effects, validity, teacher decision-making in terms of accommodations, accommodations effects for students from grades K-postsecondary education, and across five different subject areas. In addition to more studies on these topics, future research should investigate the positive effects of field-testing potential test items in accommodated formats in addition to standard formats.

Although the diversity of studies about accommodations presents a challenge to policymakers who may wish to have definitive conclusions about accommodations, the breadth of studies reflects (at least to some extent) the variety of issues present in education today. Students with disabilities are not a homogeneous group. Likewise, one accommodation does not fit all students, especially students at different grade or educational levels.

The move toward more universally accessible assessments provides an opportunity to minimize the need for accommodations. Thompson, Johnstone, and Thurlow (2002), however, noted that flexible, universally designed assessments may only minimize, not completely diminish, the need for accommodations. Likewise, until there is an individualized system for validly choosing accommodations there is a continued need for research on teacher decision-making related to accommodations.

Although accommodations research has been a part of educational research for decades, it appears that it is still in its nascence. There is still much scientific disagreement on the effects, validity, and decision-making surrounding accommodations. Such challenges lead to difficult decisions for future accommodations research. Scientific studies with large sample sizes hold promise for determining the exact effect that researchers can derive from particular accommodations. Likewise, tests that are norm-referenced and statistically defensible (such as standardized tests) lead to claims of effects on items that are more significant.

Unfortunately, much is lost on studies such as those presented above. Students with disabilities are a heterogeneous group that may require a wide variety of accommodations in order to access tests. Research from 2002-2004 was most focused on oral administration and extended time conditions for students with learning disabilities. Research related to issues for students with learning disabilities is meaningful (given the large numbers of students with learning disabilities), but research should not excessively focus on the needs of the most populous disability group. Rather, research on students with a wide variety of disabilities receiving a wide variety of accommodations is also a valuable focus, even when statistical claims are more difficult to generate.

As testing technology emerges in the 21<sup>st</sup> century, further research will need to address flexible tests that allow on-demand accommodations. Findings from 2002-2004 demonstrate only that questions still abound concerning accommodations, and that answers are often circumstantial, population-dependent, or constrained to particular tests. Such findings justify the need for tests that diminish the need for accommodations, but also more research on accommodations that currently exist. As we move forward into the next generation of accommodations research, one fact is certain: so long as policies (such as the Individuals with Disabilities Education Act and the Americans with Disabilities Act) require that students with disabilities receive accommodations, there will always be a need for research on how to best, most fairly, and most accurately assess *all* students.

## References

---

- Barton, K. E. (2002). Stability of constructs across groups of students with different disabilities on a reading assessment under standard and accommodated administrations (Doctoral dissertation, University of South Carolina, 2001). *Dissertation Abstracts International*, 62/12, 4136.
- Barton, K. E., & Huynh, H. (2003). Patterns of errors made by students with disabilities on a reading test with oral reading administration. *Educational and Psychological Measurement*, 63(4), 602-614.
- Barton, K. E., & Sheinker, A. (2003). *Comparability and accessibility: On line versus on paper writing prompt administration and scoring across students with various abilities*. Monterey, CA: CTB-McGraw-Hill.
- Bielinski, J., Ysseldyke, J., Bolt, S., Friedebach, J., & Friedebach, M. (2001). *Read-aloud accommodation: Effects on multiple-choice reading & math items*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Bielinski, J., Sheinker, A., & Ysseldyke, J. (2003). *Varied opinions on how to report accommodated test scores: Findings based on CTB/McGraw-Hill's framework for classifying accommodations* (Synthesis Report 49). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Bolt, S., & Bielinski, J. (2002). *The effects of the read aloud accommodation on math test items*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Bridgeman, B., Cline, F., & Hessinger, J. (2004). Effect of extra time on verbal and quantitative GRE scores. *Applied Measurement in Education*, 17(1), 25-37.
- Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2003). Effects of screen size, screen resolution and display rate on computer-based test performance. *Applied Measurement in Education*, 16(3), 191-205.
- Buehler, K. L. (2002). Standardized group achievement tests and the accommodation of additional time (Doctoral dissertation, Indiana State University, 2001). *Dissertation Abstracts International*, 63/04, 1312.
- Burch, M. (2002). Effects of computer-based test accommodations on the math problem-solving performance of students with and without disabilities (Doctoral dissertation, Vanderbilt University, 2002). *Dissertation Abstracts International*, 63/03, 902.

Cahalan, C., Mandinach E., & Camara, W. J. (2002). *Predictive validity of SAT I: Reasoning test for test-takers with learning disabilities and extended time accommodations*. New York, NY: The College Reporting Board.

Choi, S. W., & Tinker, T. (2002). *Evaluating comparability of paper-and-pencil and computer-based assessment in a K-12 setting*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Cisar, C. A. (2004). Teacher's knowledge about accommodations and modifications as they relate to assessment (Doctoral dissertation, Loyola University of Chicago, 2004). *Dissertation Abstracts International*, 65/10, 3754.

Crawford, L., Helwig, R., & Tindal, G. (2004). Writing performance assessments: How important is extended time? *Journal of Learning Disabilities*, 37(2), 132-142.

Dempsey, K. M. (2004). The impact of additional time on LSAT scores: Does time really matter? The efficacy of making decisions on a case-by-case basis (Doctoral dissertation, La Salle University, 2004). *Dissertation Abstracts International*, 64/10, 5212.

Elbaum, B., Arguelles, M. E., Campbell, Y., & Saleh, M. B. (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., & Marquart, A. M. (2003). *Extended time as an accommodation on a standardized mathematics test: An investigation of its effects on scores and perceived consequences for students with varying mathematical skills*. Madison, WI: University of Wisconsin- Madison, Wisconsin Center for Education Research.

Elliott, S. N., & Marquart, A. M. (2004). Extended time as a testing accommodation: Its effects and perceived consequences. *Exceptional Children*, 70(3), 349-367.

Gagnon, J. C., & McLaughlin, M. J. (2004). Curriculum, assessment, and accountability in day treatment and residential schools. *Exceptional Children*, 70(3), 263-283.

Hall, S. E. H. (2002). The impact of test accommodations on the performance of students with disabilities (Doctoral dissertation, The George Washington University, 2002). *Dissertation Abstracts International*, 63/03, 902.

Hansen, E. G., Lee, M. J., & Forer, D. C. (2002). A 'self-voicing' test for individuals with visual impairments. *Journal of Visual Impairment and Blindness*, 96(4), 273-275.

Helwig, R., Rozek-Tedesco, M. A., & Tindal, G. (2002). An oral versus a standard administration of a large-scale mathematics test. *The Journal of Special Education, 36*(1), 39-47.

Helwig, R., & Tindal, G. (2003). An Experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children, 69*(2), 211-225.

Huynh, H., Meyer, J. P., & Gallant-Taylor, D. (2002). *Comparability of scores of accommodated and non-accommodated testings for a high school exit examination of mathematics*. Paper presented at the annual meeting of the National Council on Measurement in Education. New Orleans, LA.

Huynh, H., Meyer, J. P., & Gallant, D. J. (2004). Comparability of student performance between regular and oral administrations for a high-stakes mathematics test. *Applied Measurement in Education, 17*(1), 39-57.

Idstein, B. E. (2003). Dictionary use during reading comprehension tests: An aid or a diversion? (Doctoral dissertation, Indiana University of Pennsylvania, 2003). *Dissertation Abstracts International, 64*/02, 483.

Jackson, L. M. (2003). The effects of testing adaptations on students' standardized test scores for students with visual impairments in Arizona (Doctoral dissertation, University of Arizona, 2003). *Dissertation Abstracts International, 64*/10, 3644.

Janson, I. B. (2002). The effects of testing accommodations on students' standardized test scores in a northeast Tennessee school system (Doctoral dissertation, East Tennessee State University, 2002). *Dissertation Abstracts International, 63*/02, 557.

Kappel, A. (2002). The effects of testing accommodations on subtypes of students with learning disabilities (Doctoral dissertation, University of Pittsburgh, 2002). *Dissertation Abstracts International, 63*/05, 1804.

Katzman, L. I. (2004). Students with disabilities and high stakes testing: What can the students tell us? (Doctoral dissertation, Harvard University, 2004). *Dissertation Abstracts International, 65*/05, 1732.

Ketterlin-Geller, L. R. (2003). *Establishing a validity argument for universally designed assessments*. Unpublished Doctoral Dissertation, University of Oregon, Eugene, OR.

Kettler, R. J., Niebling, B. C., Mroch, A. A., Feldman, E. S., & Newell, M. L. (2003). *Effects of testing accommodations on math and reading scores: An experimental analysis of the performance of fourth and eighth grade students with and without disabilities*. Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research.



- Kobrin, J. L., & Young, J. W. (2003). The cognitive equivalence of reading comprehension test items via computerized and paper-and-pencil administration. *Applied Measurement in Education, 16*(2), 115-140.
- Landau, S., Russell, M., Gourgey, K., Erin, J. N., & Cowan, J. (2003). Use of talking tactile tablet in mathematics testing. *Journal of Visual Impairment and Blindness, 97*(2), 85-96.
- MacArthur, C. A., & Cavalier, A. R. (2004). Dictation and speech recognition technology as test accommodations. *Exceptional Children, 71*(1), 43-58.
- McKevitt, B. C., & Elliot, S. N. (2003). Effects and perceived consequences of using read aloud and teacher-recommended testing accommodations on a reading achievement test. *The School Psychology Review, 32*(4), 583-600.
- Meloy, L. L., Deville, C., & Frisbie, D. (2002). The effect of a read aloud accommodation on test scores of students with and without a learning disability in reading. *Remedial and Special Education, 23*(4), 248-255.
- Nickerson, B. (2004). English language learners, the Stanford Achievement Test, and perceptions regarding the effectiveness of testing accommodations: A study of eighth graders (Doctoral dissertation, The George Washington University, 2004). *Dissertation Abstracts International, 65*/07, 2465.
- Pomplun, M., Frey, S., & Becker, D. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement, 62*(2), 337-354.
- Reed, E. (2002). Wrong for the right reasons: Appropriate accommodations for students with learning disabilities and/or attention deficit/hyperactivity disorder (Doctoral dissertation, Stanford University, 2002). *Dissertation Abstracts International, 63*/10, 3475.
- Scheuneman, J. D., Camara, W. J., Cascallar, A. S., Wendler, C., & Lawrence, I. (2002). Calculator access, use, and type in relation to performance in the SAT I: Reasoning test in mathematics. *Applied Measurement in Education, 15*(1), 95-112.
- Shriner, J. G., & DeStefano, L. (2003). Participation and accommodation in state assessment: The role of individualized education programs. *Exceptional Children, 69*(2), 147-161.
- Sireci, S. G., Li, S., & Scarpati, S. (2003). *The Effects of test accommodations on test performance: A review of the literature* (Research Report 485). Amherst, MA: Center for Educational Assessment.

- Sireci, S. G., Li, S., & Scarpati, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75(4), 457-490
- Tavani, C. M. (2004). The impact of testing accommodations on students with learning disabilities: An investigation of the 2000 NAEP mathematics assessment (Doctoral dissertation, The Florida State University, 2004). *Dissertation Abstracts International*, 65/07, 2493.
- Thompson, S., Blount, A., & Thurlow, M. (2002). *A summary of research on the effects of test accommodations: 1999 through 2001* (Technical Report 34). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thompson, S. J., Johnstone, C. J., Thurlow, M. L., & Altman, J. R. (2005). *2005 State special education outcomes: Five years into the decade means many changes*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thornton, A. E., Reese L. M., Pashley P. J., & Dalessandro S. P. (2002). *Predictive validity of accommodated LSAT scores*. Pennsylvania: Law School Admission Council.
- Thurlow, M. L., & Bolt, S. (2001). *Empirical support for accommodations most often allowed in state policy* (Synthesis Report No. 41). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Tindal, G., & Fuchs, L. (1999). *A summary of research on test changes: An empirical basis for defining accommodations*. Lexington, KY: University of Kentucky, Mid-South Regional Resource Center.
- Tindal, G. (2002). *Accommodating mathematics testing using a videotaped, read-aloud administration*. Washington, DC: Council of Chief State School Officers.
- Tindal, G., & Ketterlin-Geller, L. R. (2004). *Research on mathematics test accommodations relevant to NAEP testing*. Washington, DC: National Assessment Governing Board.
- Trammell, J. K. (2003). The impact of academic accommodations on final grades in a postsecondary setting. *Journal of College Reading and Learning*, 34(1), 76-90.
- Weston, T. J. (2003). *NAEP Validity Studies: The validity of oral accommodation testing*. Washington, DC: National Center for Education Statistics.

Woods, K. (2004). Deciding to provide a reader in examinations for the general certificate of secondary education (GCSE): Questions about Validity and Inclusion. *British Journal of Special Education*, 31(3), 122-124.

## Appendix A—Summary of Research Purpose

<b>Determine the effect of the use of accommodations on test scores of students with disabilities</b>	
Bolt S., & Bielinski J. (2002). <i>The effects of the read aloud accommodation on math test items</i> . Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.	The purpose of this study was to examine whether score comparability improves when a test is read aloud to students.
Buehler K. L. (2002). Standardized group achievement tests and the accommodation of additional time (Doctoral dissertation, Indiana State University, 2001). <i>Dissertation Abstracts International</i> , 63/04, 1312.	This study investigated the effects of additional time on test data.
Burch M. (2002). Effects of computer-based test accommodations on the math problem-solving performance of students with and without disabilities (Doctoral dissertation, Vanderbilt University, 2002). <i>Dissertation Abstracts International</i> , 63/03, 902.	This study investigated three different computer-based testing accommodations. Students were tested in the following conditions: Standard administration (SA), computer-read text (CRT), video (V), constructed responses (CON), and comprehensive accommodations (CA).
Crawford, L., Helwig, R., & Tindal, G. (2004). Writing performance assessments: How important is extended time? <i>Journal of Learning Disabilities</i> , 37(2), 132-142.	This study investigated the effects of varying the available amounts of testing time on the writing performance of students in general and special education at Grades 5 and 8.
Barton K. E., & Sheinker A. (2003). <i>Comparability and accessibility: On line versus on paper writing prompt administration and scoring across students with various abilities</i> . Monterey, CA: CTB-McGraw-Hill.	The purpose of this study was to examine whether students without disabilities obtain higher scores than those with disabilities when all participants were administered two writing prompts (on-line and paper-based) counter-balanced by mode of administration and prompt.
Dempsey, K. M. (2004). The impact of additional time on LSAT scores: Does time really matter? The efficacy of making decisions on a case-by-case basis (Doctoral dissertation, La Salle University, 2004). <i>Dissertation Abstracts International</i> , 64/10, 5212.	This study examines the relationship between cognitive test data and Law School Admission Test (LSAT) performance as well as the effects of being granted additional test time to take the LSAT. This study also evaluated the difference between candidates' standard and accommodated LSAT scores and their predicted LSAT scores.
Elbaum B., Arguelles M. E., Campbell Y., & Saleh M. B. (2004). Effects of a Student-Reads-Aloud Accommodation on the Performance of Students With and Without Learning Disabilities on a Test of Reading Comprehension. <i>Exceptionality</i> , 12(2), 71-87.	This study examined the impact of a student-reads-aloud (i.e., student reads text but aloud) accommodation on the performance of middle and high school students with and without learning disabilities on a test of reading comprehension.

<p>Elliott S. N., &amp; Marquart A. M. (2003). <i>Extended time as an accommodation on a standardized mathematics test: An investigation of its effects on scores and perceived consequences for students with varying mathematical skills</i>. Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research.</p>	<p>This study investigated the significance on scoring for students who took equivalent forms of a standardized math test in two conditions (extended time and standard time).</p>
<p>Elliott, S. N., &amp; Marquart, A. M. (2004). Extended time as a testing accommodation: Its effects and perceived consequences. <i>Exceptional Children</i>, 70(3), 349-367.</p>	<p>This investigation examined the effect of extended time on the performance of students with disabilities, students educationally at risk in math, and students without disabilities.</p>
<p>Hansen E. G., Lee M. J., &amp; Forer D. C. (2002). A 'self-voicing' test for individuals with visual impairments. <i>Journal of Visual Impairment and Blindness</i>, 96(4), 273-275.</p>	<p>The study investigated the use of speech output technology for tests for individuals with visual impairments.</p>
<p>Helwig R., Rozek-Tedesco M. A., &amp; Tindal G. (2002). An oral versus a standard administration of a large-scale mathematics test. <i>The Journal of Special Education</i>, 36(1), 39-47.</p>	<p>The purpose of this study was to examine whether students perform better on examinations when read aloud items via a video presentation.</p>
<p>Idstein, B. E. (2003). Dictionary use during reading comprehension tests: An aid or a diversion? (Doctoral dissertation, Indiana University of Pennsylvania, 2003). <i>Dissertation Abstracts International</i>, 64/02, 483.</p>	<p>The purpose of this study is to examine students' use of dictionaries during reading comprehension exams, especially since their use during exams has come under critical reexamination.</p>
<p>Janson I. B. (2002). The effects of testing accommodations on students' standardized test scores in a northeast Tennessee school system (Doctoral dissertation, East Tennessee State University, 2002). <i>Dissertation Abstracts International</i>, 63/02, 557.</p>	<p>Scores obtained by students who received special education services and did not receive accommodations in 1998 and/or 1999 were compared to scores obtained by the same students who did receive accommodations in later testing. Ninety-nine percent of students who received accommodations were given the read aloud accommodation.</p>
<p>Kappel A. (2002). The effects of testing accommodations on subtypes of students with learning disabilities (Doctoral dissertation, University of Pittsburgh, 2002). <i>Dissertation Abstracts International</i>, 63/05, 1804.</p>	<p>This study investigated the effects of two testing accommodations, extended time and oral administration, on the math test performance of students' with learning disabilities</p>
<p>Kettler R. J., Niebling, B. C., Mroch A. A., Feldman E. S., &amp; Newell M. L. (2003). <i>Effects of testing accommodations on math and reading scores: An experimental analysis of the performance of fourth and eighth grade students with and without disabilities</i>. Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research.</p>	<p>Participants with disabilities were assigned accommodations based on their IEPs. Participants without disabilities were paired with students with disabilities. Each pair was tested under the accommodation condition during which both students in the pair received the same set of accommodations.</p>

<p>Landau, S., Russell, Gourgey, K., Erin, J. N., &amp; Cowan, J. (2003). Use of talking tactile tablet in mathematics testing. <i>Journal of Visual Impairment and Blindness</i>, 97(2), 85-96.</p>	<p>This study examined the extent to which use of the Talking Tactile Tablet had a positive impact on the mathematics performance of students who were visually impaired and/or had difficulty visualizing graphics and diagrams. To the extent possible, the study also explored the Talking Tactile Tablet's impact on the difficulty of items.</p>
<p>Macarthur, C. A., &amp; Cavalier, A. R. (2004). Dictation and speech recognition technology as test accommodations. <i>Exceptional Children</i>, 71(1), 43-58.</p>	<p>This study addressed the feasibility and validity of dictation using speech recognition software (Dragon Naturally Speaking, Version 4) and dictation to a scribe as test accommodations for students with learning disabilities.</p>
<p>McKevitt, B. C., &amp; Elliott, S. N. (2003). Effects and perceived consequences of using read aloud and teacher-recommended testing accommodations on a reading achievement test. <i>The School Psychology Review</i>, 32(4), 583-600.</p>	<p>The purpose of this study was to test students' performance on a reading test with and without read-aloud accommodations.</p>
<p>Reed, E. (2002). Wrong for the right reasons: Appropriate accommodations for students with learning disabilities and/or attention deficit/hyperactivity disorder (Doctoral dissertation, Stanford University, 2002). <i>Dissertation Abstracts International</i>, 63/10, 3475.</p>	<p>This study considered student performance and the appropriateness of accommodations at the level of the individual student through a think-aloud process. Students were asked to think-aloud while solving grade level mathematics problems.</p>
<p>Tavani, C. M. (2004). The impact of testing accommodations on students with learning disabilities: An investigation of the 2000 NAEP mathematics assessment (Doctoral dissertation, The Florida State University, 2004). <i>Dissertation Abstracts International</i>, 65/07, 2493.</p>	<p>This study addressed the effects of accommodations on mathematical performance scores and examined additional variables that showed to have strong relationships with student's test performances.</p>
<p>Tindal G. (2002). <i>Accommodating mathematics testing using a videotaped, read-aloud administration</i>. Washington, DC: Council of Chief State School Officers.</p>	<p>Students participated in both standard and videotaped test administrations. During the videotaped administration the test items were read aloud individually, in a paced format, with visual prompting of the answer choices.</p>
<p>Trammell, J. K. (2003). The impact of academic accommodations on final grades in a postsecondary setting. <i>Journal of College Reading and Learning</i>, 34(1), 76-90.</p>	<p>The purpose of this study was to determine whether postsecondary students with learning disabilities and/or Attention Deficit Disorder experienced a differential increase in end-of-term grades when they used academic accommodations required by the Americans with Disabilities Act. Students received one or more of the following academic accommodations throughout the school year: additional time to complete the tests, taping classes, testing in a separate room and books on tape.</p>

<p>Weston, T. J. (2003). <i>NAEP validity studies: The validity of oral accommodation testing</i>. Washington, DC: National Center for Education Statistics.</p>	<p>This study examined three factors related to read aloud accommodation is math. First, accommodated test scores were compared to non-accommodated scores for a sample of students with learning disabilities. Second was the relative benefit that students with learning disabilities received from read aloud accommodations. Finally, the author examined the accuracy of information derived from accommodated and non-accommodated tests.</p>
<p><b>Investigate the effects of accommodations on test score validity</b></p>	
<p>Barton K. E. (2002). Stability of constructs across groups of students with different disabilities on a reading assessment under standard and accommodated administrations (Doctoral dissertation, University of South Carolina, 2001). <i>Dissertation Abstracts International</i>, 62/12, 4136.</p>	<p>The purpose of this study was to examine whether a similar construct is measured among students who are administered either the oral accommodation (OA) form or a regular form of an assessment.</p>
<p>Bridgeman, B., Cline, F., &amp; Hessinger, J. (2004). Effect of extra time on verbal and quantitative GRE scores. <i>Applied Measurement in Education</i>, 17(1), 25-37.</p>	<p>The purpose of this study was to examine the effects of extra time on the Graduate Record Examination General Test.</p>
<p>Bridgeman, B., Lennon, M. L., &amp; Jackenthal, A. (2003). Effects of screen size, screen resolution and display rate on computer-based test performance. <i>Applied Measurement in Education</i>, 16(3), 191-205.</p>	<p>This study evaluated the effects of variations in screen size, resolution, and presentation delay on verbal and mathematics scores. There were three screen display conditions (size and resolution) crossed with two presentation rate conditions (delay or no delay).</p>
<p>Cahalan C., Mandinach E., &amp; Camara W. J. (2002). <i>Predictive validity of SAT I: Reasoning test for test-takers with learning disabilities and extended time accommodations</i>. New York, NY: The College Reporting Board.</p>	<p>The study was conducted to examine the predictive validity of scores taken with an extended time accommodation.</p>
<p>Hall S. E. H. (2002). The impact of test accommodations on the performance of students with disabilities (Doctoral dissertation, The George Washington University, 2002). <i>Dissertation Abstracts International</i>, 63/03, 902.</p>	<p>Subjects received a variety of accommodations including extended time, dictated response, small group, and oral administration of the test.</p>
<p>Huynh H., Meyer J. P., &amp; Gallant-Taylor D. (2002). <i>Comparability of scores of accommodated and non-accommodated testings for a high school exit examination of mathematics</i>. Paper presented at the annual meeting of the National Council on Measurement in Education. New Orleans, LA.</p>	<p>Students received a different form of the test that was designed to be appropriate for testing students with visual and hearing impairments. This form could have been provided in a regular print, large-print, or loose-leaf version. This form also may have been administered orally or by sign language to some of the students.</p>

Huynh, H., Meyer, J. P., & Gallant, D. J. (2004). Comparability of student performance between regular and oral administrations for a high-stakes mathematics Test. <i>Applied Measurement in Education</i> , 17(1), 39-57.	This study examined the effect of oral administration accommodations on test structure and student performance on the mathematics portion on the South Carolina high School Exit Examination.
Kobrin, J. L., & Young, J. W. (2003). The cognitive equivalence of reading comprehension test items via computerized and paper-and-pencil administration. <i>Applied Measurement in Education</i> , 16(2), 115-140.	The cognitive equivalence of computerized and paper-and-pencil reading comprehension tests was investigated.
Meloy L. L., Deville C., & Frisbie D. (2002). The effect of a read aloud accommodation on test scores of students with and without a learning disability in reading. <i>Remedial and Special Education</i> , 23 (4), 248-255.	Students were randomly assigned to two experimental conditions. In one condition the test was administered according to standard procedures; in the other condition the test was read aloud to the students.
Pomplun M., Frey S., & Becker D. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. <i>Educational and Psychological Measurement</i> , 62(2), 337-354.	Students took two forms of a test in computerized and paper-and-pencil versions.
Scheuneman J. D., Camara W. J., Cascallar A. S., Wendler C., & Lawrence I. (2002). Calculator access, use, and type in relation to performance in the SAT I: Reasoning test in mathematics. <i>Applied Measurement in Education</i> , 15(1), 95-112.	After completing the test, participants were asked to respond to a set of three questions about their use of a calculator during the test.
Thornton A. E., Reese L. M., Pashley P. J., & Dalessandro S. P. (2002). <i>Predictive validity of accommodated LSAT scores</i> . Pennsylvania: Law School Admission Council.	The validity of scores obtained by test takers who were administered the test under nonstandard time conditions (i.e., accommodations that included extended time) was investigated.
Woods, K. (2004). Deciding to provide a reader in examinations for the General Certificate of Secondary Education (GCSE): Questions about validity and inclusion. <i>British Journal of Special Education</i> , 31(3), 122-124.	This study examined the effects of providing a read-aloud accommodated test to examinees in England, Wales and Northern Ireland. Its purpose was to report whether reading age and self-prediction were accurate indicators of the need for read-aloud accommodations.
<b>Study institutional factors, teacher judgment, or student desirability of accommodation use</b>	
Bolt, S. E. (2004). <i>Using DIF analyses to examine several commonly-held beliefs about testing accommodations for students with disabilities</i> . Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.	The purpose of this study was to examine the extent of data based support for several commonly held opinions about testing accommodations for students with disabilities.



<p>Bielinski, J., Sheinker, A., &amp; Ysseldyke, J. (2003). <i>Varied opinions on how to report accommodated test scores: Findings based on CTB/McGraw-Hill's framework for classifying accommodations</i> (Synthesis Report 49). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.</p>	<p>A list of 44 different accommodations categorized into presentation, response, setting, and timing accommodations were used.</p>
<p>Cisar, C. A. (2004). Teacher's knowledge about accommodations and modifications as they relate to assessment (Doctoral dissertation, Loyola University of Chicago, 2004). <i>Dissertation Abstracts International</i>, 65/10, 3754.</p>	<p>The purpose of this study was to determine if there was a difference among staff (administrators, general education teachers, special education teachers, elective area teachers: art, PE, music) servicing students with special needs in their ability to distinguish accommodations from modifications and their ability to use them in assessment activities.</p>
<p>Gagnon, J. C., &amp; McLaughlin, M. J. (2004). Curriculum, assessment, and accountability in day treatment and residential schools. <i>Exceptional Children</i>, 70(3), 263-283.</p>	<p>This study determined school-level curricular, assessment, and accountability policies and practices in private and public day treatment and residential schools for elementary age children with emotional or behavioral disorders.</p>
<p>Helwig, R., &amp; Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. <i>Exceptional Children</i>, 69(2), 211-225.</p>	<p>This study tested the accuracy with which special education teachers determine which students need read-aloud accommodations. An additional goal of this study was to develop a profile of students who benefit from this type of accommodation by contrasting their achievement levels in reading and basic math skills.</p>
<p>Jackson, L. M. (2003). The effects of testing adaptations on students' standardized test scores for students with visual impairments in Arizona (Doctoral dissertation, University of Arizona, 2003). <i>Dissertation Abstracts International</i>, 64/10, 3644.</p>	<p>The purpose of this study was to determine the relationship of testing modifications, a type of adaptation, and the effects of demographic information on students' standardized test scores for students in Arizona who have visual impairments including those with additional disabilities.</p>
<p>Katzman, L. I. (2004). Students with disabilities and high stakes testing: What can the students tell us? (Doctoral dissertation, Harvard University, 2004). <i>Dissertation Abstracts International</i>, 65/05, 1732.</p>	<p>This study examined the qualitative aspects of high stakes testing and accommodations for students with disabilities by asking students to explain their understanding and experiences of participating in a large-scale high school examination.</p>
<p>Nickerson, B. (2004). English language learners, the Stanford Achievement Test, and perceptions regarding the effectiveness of testing accommodations: A study of eighth graders (Doctoral dissertation, The George Washington University, 2004). <i>Dissertation Abstracts International</i>, 65/07, 2465.</p>	<p>This study examined the perception of students with regard to the effectiveness of testing accommodations in assisting them to more accurately demonstrate content knowledge and skills.</p>

<p>Shriner, J. G., &amp; Destefano, L. (2003). Participation and accommodation in state assessment: The role of individualized education programs. <i>Exceptional Children</i>, 69(2), 147-161.</p>	<p>The purpose of this study was to test if training sessions help special education teachers and administrators use, and report accommodations on test day.</p>
<p><b>Examine patterns of errors across items or tests</b></p>	
<p>Barton, K. E., &amp; Huynh, H. (2003). Patterns of errors made by students with disabilities on a reading test with oral reading administration. <i>Educational and Psychological Measurement</i>, 63(4), 602-614.</p>	<p>This study examined differences in the types of errors made by students with disabilities on a multiple choice reading test administered under oral reading accommodations.</p>
<p>Choi, S. W., &amp; Tinker T. (2002). <i>Evaluating comparability of paper-and-pencil and computer-based assessment in a K-12 setting</i>. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.</p>	<p>Students took alternate forms of the test under computer administration and paper-and-pencil administration.</p>
<p><b>Meta-analysis</b></p>	
<p>Sireci, S. G., Li, S., &amp; Scarpati, S. (2003). <i>The effects of test accommodations on test performance: A review of the literature</i> (Research Report 485). Amherst, MA: Center for Educational Assessment.</p>	<p>The purpose of this study was to analyze existing research in the area of test accommodations.</p>
<p>Tindal, G., &amp; Ketterlin-Geller, L.R. (2004). <i>Research on mathematics test accommodations relevant to NAEP testing</i>. Washington, DC: National Assessment Governing Board.</p>	<p>The purpose of this study was to synthesize research pertaining to differential item functioning.</p>



## Appendix B—Summary of Type of Assessment

Author	Norm-Referenced and Other Standardized Tests	State Criterion Referenced Tests of Performance Assessments	School or District-designed Tests	Other
Barton (2002)		The reading portion of a secondary level statewide assessment was used as the dependent variable.		
Barton (2003)		The study was based on statewide data from the 1996, 1997, and 1998 administration of the Oral Accommodation form of South Carolina's statewide reading test.		
Barton (2003)				Questionnaires addressing computer literacy, accommodations, and accessibility were given to all students participating. Writing prompts were scored via paper and online formats and analyzed via descriptive and inferential analysis.

<b>Bielinski (2003)</b>				<p>This study is a survey of the perceptions held by people familiar with policy or research on the way in which test scores are influenced by accommodations and how scores obtained under accommodated conditions are to be treated in reporting. Participants marked each accommodation as either: 1) measuring the construct in the same way, 2) changing the meaning of the test score, or 3) not having definitive evidence to place it in either category one or two.</p>
<b>Bolt (2002)</b>		<p>Level analyses were conducted on multiple choice math items from the Missouri Assessment Program.</p>		
<b>Bolt (2004)</b>		<p>A series of DIF analyses were conducted across three statewide achievement tests.</p>		

<b>Bridgeman (2003)</b>	The participants were tested using a computerized version of questions from the SAT 1: Reasoning Test. The test was given in various formats crossing screen size, resolution, and presentation rate.			
<b>Bridgeman (2004)</b>	The verbal and quantitative sections of the Graduate Record Examination were experimentally administered with standard time limits and at 1.5 times the normal allotted time.			
<b>Buehler (2002)</b>	Students were administered the reading subtests of the California Achievement Tests, Fifth Edition (CAT/5) and the rapid-naming subtests of the Comprehensive Test of Phonological Processing (CTOPP).			
<b>Burch (2002)</b>				Several math tests were administered to assess the math problem-solving performance of students with and without disabilities.

<b>Cahalan (2002)</b>	SAT I test scores and self-reported high school grade point average (HSGPA) to predict first year grade point average (FGPA).			
<b>Choi (2002)</b>		Items from a statewide math and reading test were analyzed. Surveys were also conducted to determine student computer experience.		
<b>Cisar (2004)</b>				The researcher created a questionnaire that contained four sections of data collection: professional development, identification, accommodation information, and demographics.
<b>Crawford (2004)</b>		The students completed a 30-minute Oregon state writing performance assessment as well as a longer writing performance assessment which was completed over 3 days. Assessments were evaluated on four traits (ideas, organization, conventions, and sentence fluency).		

<b>Dempsey (2003)</b>	Subjects were administered the LSAT under extended time conditions.			
<b>Elbaum (2004)</b>		A test made up of 3 <sup>rd</sup> –5 <sup>th</sup> grade reading passages.		
<b>Elliott (2003)</b>	Students completed alternate short forms of standardized mathematics tests developed from the TerraNova Level 18 mathematics test.			
<b>Elliott (2004)</b>	The students completed one of two alternate short forms of standardized mathematics tests developed from the TerraNova level 18 mathematics test. Upon completion of the test, students completed an accommodations survey about their reactions to working on the test under the accommodated or standard conditions.			



<b>Gagnon (2004)</b>				Two surveys were developed based on a review of literature, consideration of current educational reform, etc. There were five sections on each survey, however, this study focuses only on the sections on curricular policies and accountabilities.
<b>Hall (2002)</b>		Results from the 2000 administration of the Maryland School Performance Assessment Program (MSAP) were used as the dependent variable in this post hoc analysis.		
<b>Hansen (2002)</b>				The study examined the use of a prototype testing system that utilizes synthesized speech to deliver questions on reading and listening comprehension tests.

<b>Helwig (2002)</b>		Items were selected from a statewide multiple-choice math test; items considered more difficult to read were specifically analyzed.		
<b>Helwig (2003)</b>				For each student, the appropriate teacher completed a survey that rated the student's skill level in both reading and mathematics on a 5-point Likert scale. The teacher also predicted which students would benefit most from a read-aloud accommodation. The students were also tested with a standardized reading and basic math skills test.
<b>Huynh (2002)</b>		Performance on the math portion of the South Carolina High School Exit Examination was used as the dependent variable.		

<b>Huynh (2004)</b>		The mathematics section of the state exit examination was given at grade 10. All students in this assessment program were given unlimited time to complete the test.		
<b>Idstein (2003)</b>		The students were given a test called a bagrut, a reading comprehension test in English. The reading task component included two reading passages. Students were permitted use of the <i>Oxford Student's Dictionary for Hebrew Speakers</i> .		
<b>Jackson (2003)</b>		The dependant variable in this study was the Stanford Achievement Test, 9 <sup>th</sup> Edition.		
<b>Janson (2002)</b>		Performance on the Tennessee Comprehensive Assessment Program (TCAP) achievement test was used as the dependent variable. Results from the following years were analyzed: 1998, 1999, 2000, and 2001.		

<b>Kappel (2002)</b>		All students were administered items from the Mathematics subtest of the California Achievement Test under several conditions, with and without accommodations.		
<b>Katzman (2004)</b>				The students were interviewed after completing the 10 <sup>th</sup> grade MCAS, an assessment students must pass in order to graduate.
<b>Kettler (2003)</b>	Two math subtests and two reading subtests from research editions of the TerraNova Multiple Assessment Battery were used to assess participants' achievement levels.			
<b>Kobrin (2003)</b>	Subjects were tested using reading comprehension items from the ETS GRE General Test Big Book. Two long passages consisting of 55 lines and 7 corresponding test items were selected and administered via computer and paper-and pencil formats.			

<p><b>Landau (2003)</b></p>				<p>For this study, three mathematics test forms, each containing four items, were administered to the participants. Each of the 12 items referenced a diagram or graphical element. The items focused on geometry, measurement, patterns and relations, and statistics and probability.</p>
<p><b>Macarthur (2004)</b></p>	<p>Two measures were used to evaluate accuracy of speech recognition: sentence probes and word-list probes. Students wrote essays under the following three conditions: using handwriting, using a scribe, and using speech recognition software.</p>			

<p><b>McKevitt (2003)</b></p>	<p>Two forms of a research version of the TerraNova Multiple Assessments Reading test (eighth-grade level) were used in this study. After completing the test, students completed a survey about testing accommodations. Teachers also completed a survey about their perceptions of the effectiveness of testing with accommodations.</p>			
<p><b>Meloy (2002)</b></p>		<p>Participants were administered four tests from the Iowa Tests of Basic Skills (ITBS): Science, Usage and Expression, Math Problem-Solving, Data Interpretation, and Reading Comprehension.</p>		
<p><b>Nickerson (2004)</b></p>				<p>Interviews were used to elicit student perceptions about which testing accommodations effectively assisted the students in demonstrating what he/she knows and can do on the SAT-9 test.</p>

<b>Pomplun (2002)</b>	Participants were administered multiple forms of a reading placement test, namely the Nelson-Denny Reading Test.			
<b>Reed (2002)</b>	Three subtests of the Woodcock-Johnson Revised Tests and a subset of tasks from the Wechsler Intelligence Scale for Children were used in this study. The think-aloud data were analyzed in conjunction with pre-test ability and achievement measures to determine why students got test items right or wrong. Test items were thought to function appropriately for students who got the answer wrong and displayed no mastery of the construct being assessed during the think-aloud process.			

<p><b>Scheuneman (2002)</b></p>	<p>The participants were administered the SAT I: Reasoning test in Mathematics in domestic test centers. Questions about use of the calculator on the test were placed in the answer sheets for the November 1996 and the November 1997 administrations of the examination.</p>			
<p><b>Shriner (2003)</b></p>		<p>The IEP analyses that are reported were conducted twice (1999, 2000). During the intervening year, site-based management teams of special and general education teachers and administrators participated in a series of training sessions and follow-up conducted by the researchers during March 1999 through February 2000. The decisions of the same groups of trained IEP team members were followed through both years in a longitudinal design.</p>		



<b>Tavani (2004)</b>	The 2000 NAEP Mathematics assessment.			
<b>Thornton (2002)</b>	The measure used to assess the predictive validity of the LSAT for participant groups was law school first year average grades.			
<b>Tindal (2002)</b>		Fourth and seventh grade levels of a multiple choice mathematics test were administered in two different forms (videotaped and standard) in a counterbalanced order.		
<b>Trammell (2003)</b>				End of term grades for each subject were compared and contrasted in this study.
<b>Weston (2003)</b>	All subjects took two matched forms of a mathematics assessment based on NAEP items: one form accommodated (read-aloud), and one form non-accommodated. All students also took the first part of the Third Grade TerraNova Reading test to determine reading level.			

<b>Woods (2004)</b>	The GCSE is a high stakes examination in England, Wales and Northern Ireland. It is used as a predictor of future educational achievement.			
---------------------	--	--	--	--



## Appendix C— Subject Area Studied (by Author)

Author	Math	Reading/ Language Arts	Science	Writing	Social Studies	No Specific Content Area	Total
Barton (2002)		X					1
Barton (2003)		X					1
Barton (2003)				X			1
Bielinski (2003)							N/A
Bolt (2002)		X					1
Bolt (2004)	X	X					2
Bridgeman (2003)	X	X					2
Bridgeman (2004)	X	X					2
Buehler (2002)		X					1
Burch (2002)	X	X					2
Cahalan (2002)						X	1
Choi (2002)	X	X					2
Cisar (2004)							N/A
Crawford (2004)				X			1
Dempsey (2003)						X	1
Elbaum (2004)		X					1
Elliott (2003)	X						1
Elliott (2004)	X						1
Gagnon (2004)							N/A
Hall (2002)		X	X				2
Hansen (2002)		X					1
Helwig (2002)		X					1
Helwig (2003)	X	X					2
Huynh (2002)	X						1
Huynh (2004)				X			1
Idstein (2003)		X					1
Jackson (2003)	X	X					2
Janson (2002)	X		X		X		3
Kappel (2002)	X						1
Katzman (2004)							N/A
Kettler (2003)	X	X					2
Kobrin (2003)		X					1
Landau (2003)	X						1
Macarthur (2004)	X						1
McKevitt (2003)		X					1

Meloy (2002)	X	X	X				3
Nickerson (2004)							N/A
Pomplun (2002)		X					1
Reed (2002)	X	X					2
Scheuneman (2002)	X						1
Shriner (2003)						X	1
Tavani (2004)	X						1
Thornton (2002)						X	1
Tindal (2002)	X						1
Trammell (2003)						X	1
Weston (2003)	X	X					2
Woods (2004)						X	1

**Appendix D —Type of Accommodation Studied (by Author)** 

Author	Presentation				Response			Setting		Timing		Techno-logical Aid		Other
	Oral Presentation	Computer Administration	Dictionary Use	Large Print	Dictated Response	Word Processor	Calculator	Individual/ Small Group Setting	Extended Time	Multiple Day	Video/ Techno. Aid	Multiple		
Barton (2002)	X													
Barton (2003)	X													
Barton (2003)		X										X		
Bielinski (2003)														
Bolt (2002)	X													
Bolt (2004)												X		
Bridgeman (2003)		X							X					
Bridgeman (2004)									X					
Buehler (2002)														
Burch (2002)														
Cahalan (2002)									X					
Choi (2002)		X												
Cisar (2004)														N/A
Crawford (2004)										X				
Dempsey (2003)									X					
Elbaum (2004)														X
Elliott (2003)									X					
Elliott (2004)									X					
Gagnon (2004)														N/A
Hall (2002)												X		
Hansen (2002)														
Helwig (2002)	X													
Helwig (2003)														X
Huynh (2002)														X
Huynh (2004)	X													
Idstein (2003)			X											







## Appendix E—Summary of Participants

Author	Number of Study Participants and Percent with Disabilities	Grade-Level of Participants	Types of Disabilities of Students Included in the Sample (as labeled by authors)
Barton (2002)	5,921 (28% students with disabilities)	10 <sup>th</sup> , 12 <sup>th</sup> Grades	Learning disability, emotional disability, mental retardation, speech, language, vision or hearing impairment, physical disability
Barton (2003)	2,924 (80% students with disabilities.)	12 <sup>th</sup> Grade	Learning disability, mentally challenged, emotional disability, physical disability, communication disability
Barton (2003)	630 (50% students with disabilities)	4 <sup>th</sup> –6 <sup>th</sup> Grades	Emotionally disturbed, learning disabled, physically disabled, speech/language, hearing impaired
Bielinski (2003)	86 (% N/A)	State assessment directors, state special education directors, or individuals who have printed research on test accommodations or have published accommodations research.	N/A
Bolt (2004)	More than 1,000 (Number of students with disabilities not reported)	Elementary School, High School	Not Reported
Bolt, Bielinski J (2002)	3,013 (67% students with disabilities)	4 <sup>th</sup> Grade.	
Bridgeman (2003)	357 (Number of students with disabilities not reported)	11 <sup>th</sup> Grade	Not Reported
Bridgeman (2004)	7,653	Post secondary students	Not reported
Buehler (2002)	49 (45% students with disabilities)	K-5 <sup>th</sup> Grades	Not reported
Burch (2002)	49 (67% students with disabilities)	4 <sup>th</sup> Grade	Reading deficit
Cahalan (2002)	34,000 (Number of students with disabilities not reported)	College students	Not reported
Choi (2002)	1600 (Number of students with disabilities not reported)	3 <sup>rd</sup> , 10 <sup>th</sup> Grades	Not reported
Cisar (2004)	505 (% N/A)	School staff members	N/A
Crawford (2004)	353 (14% students with disabilities)	5 <sup>th</sup> –8 <sup>th</sup> Grades	Not reported

Dempsey (2003)	200	Post collegiate adults	Participants reported either attention deficit or a learning disability limiting performance under standard test conditions
Elbaum (2004)	311 (74% students with disabilities)	6 <sup>th</sup> –10 <sup>th</sup> grade students	Not Reported
Elliott (2003)	69 (33% students with disabilities)	8 <sup>th</sup> Grade	Mild learning disabilities, emotional disabilities, behavioral disabilities, mild physical disabilities, speech and language disabilities, mild cognitive disabilities
Elliott (2004)	97 (24% students with disabilities)	8 <sup>th</sup> Grade	Mild learning disabilities, emotional disabilities, behavioral disabilities, mild physical disabilities, speech and language disabilities, mild cognitive disabilities
Gagnon (2004)	500 (% N/A)	Principals and teachers	N/A
Hall (2002)	192,000 (6% students with disabilities)	5 <sup>th</sup> Grade	Not reported
Hansen (2002)	17 (100% students with disabilities)	Ages 17 to 55	Legally blind
Helwig (2002)	1,343 (20% students with disabilities)	4 <sup>th</sup> , 5 <sup>th</sup> , 7 <sup>th</sup> , and 8 <sup>th</sup> Grades	Reading
Helwig (2003)	1,218 (20% students with disabilities)	4 <sup>th</sup> –8 <sup>th</sup> Grades	Learning disability, language impairment, serious emotional disturbance, mental retardation
Huynh (2002)	90,000 (8% students with disabilities)	10 <sup>th</sup> Grade	Speech, hearing, visual, orthopedic, emotional, learning disabilities, educable mentally retarded, and trainable mentally retarded
Huynh (2004)	89,214 (4% students with disabilities).	10 <sup>th</sup> Grade	Speech, hearing, visual, orthopedic, emotional, learning disabilities, educable mental retardation, and trainable mental retardation
Idstein (2003)	63 (Number of students with disabilities not reported)	11 <sup>th</sup> Grade	Not reported

Jackson (2003)	71 (100% students with disabilities)	2 <sup>nd</sup> –9 <sup>th</sup> Grades	Visual impairments, including students with additional disabilities. The students in this study attended either a specialized school for the visually impaired or a public school with support from teachers of the visually impaired
Janson (2002)	448 (100% students with disabilities)	2 <sup>nd</sup> –8 <sup>th</sup> Grades	Twelve disability groups represented
Kappel (2002)	47 (100% students with disabilities).	5 <sup>th</sup> grade	Students with learning disabilities were categorized into one of three groups based on patterns of performance on a large-scale achievement test, which was administered to all participants. A fourth group of students without disabilities was also included in the sample.
Katzman (2004)	36 (67% students with disabilities)	10 <sup>th</sup> Grade	Not Reported
Kettler (2003)	196 (44% students with disabilities)	4 <sup>th</sup> , 8 <sup>th</sup> Grades	Not reported
Kobrin (2003)	48 (Number of students with disabilities not reported)	College students	Not reported
Landau (2003)	8 (100% students with disabilities)	9 <sup>th</sup> Grade-College students	Visual impairments resulting in a need for braille
Macarthur (2004)	31 (68% students with disabilities)	High school students	Not reported
McKevitt (2003)	79 (51% students with disabilities)	8 <sup>th</sup> Grade	Not reported
Meloy (2002)	260 (24% students with disabilities)	6 <sup>th</sup> –8 <sup>th</sup> grades	Reading deficit
Nickerson (2004)	30 (Number of students with disabilities not reported)	8 <sup>th</sup> Grade	Not Reported
Pomplun (2002)	215 (Number of students with disabilities not reported)	High school, post-secondary	Not reported
Reed (2002)	36 (78% students with disabilities)	8 <sup>th</sup> Grade	Learning disability, attention disability
Scheuneman (2002)	417,000 (Number of students with disabilities not reported)	11 <sup>th</sup> , 12 <sup>th</sup> Grades	Not reported
Shriner (2003)	651 (92% students with disabilities)	3 <sup>rd</sup> –11 <sup>th</sup> Grades	Learning disability, behavior disorder, mental retardation, speech/language, orthopedic impairment, visual impairment, hearing impairment, autism, other health impairment

Sireci (2003)	N/A	N/A	N/A
Tavani (2004)	42,453 (5% students with disabilities)	4 <sup>th</sup> Grade, 8 <sup>th</sup> Grade, 12 <sup>th</sup> Grade	Not Reported
Thornton (2002)	123,065 (1% students with disabilities)	Law school students	Attention deficit, learning disability, neurological impairment, and visual impairment subgroups.
Tindal (2002)	2,000 (40% students with disabilities)	4 <sup>th</sup> , 5 <sup>th</sup> , 7 <sup>th</sup> , and 8 <sup>th</sup> Grades	Mental, speech, orthopedic, traumatic, learning disability, hearing, visual, autism
Tindal (2004)	N/A	N/A	N/A
Trammell (2003)	61 (100% students with disabilities)	Undergraduate college students	Learning disability, attention deficit
Weston (2003)	119 (54% students with disabilities)	4 <sup>th</sup> Grade	Not reported
Woods (2004)	38 (Number of students with disabilities not reported)	High school students	Not reported

## Appendix F—Summary of Research Results

Author	
Barton (2002)	The results indicate that a similar construct was measured among students with and without disabilities taking the regular form. The results also indicate that a similar construct was measured among students with and without disabilities taking the oral accommodation form.
Barton (2003)	The study indicates that when errors are used as an extra factor in exploring the nature of proficiency, the reading construct varies only slightly across disability groups. The results indicate that it is safe to apply the same meaning to test scores for these groups even when the test is administered under different accommodations.
Barton (2003)	Results indicated that students without disabilities obtain higher scores than those with disabilities. There were significant differences between essays scored online and by hand scorers; however, there were no differences between students' performance online or on paper.
Bielinski (2003)	The results show that the extent of agreement about how accommodated scores should be treated depends on the accommodation. The study also shows how deep-seated beliefs lead some respondents to consider almost no accommodation as changing the construct, whereas other respondents consider almost all accommodations as influencing the construct being measured.
Bolt (2002)	The read aloud accommodation did not appear to improve score comparability for students with reading disabilities when compared to students without disabilities. More items displaying differential item functioning (DIF) were identified for those who received the accommodation than for those who did not receive the accommodation.
Bolt (2004)	Results provide some support for the commonly held beliefs, although results were not always consistent across datasets. The results also point to the challenge of appropriately assessing the skills and knowledge of students with disabilities using currently available assessments.
Bridgeman (2003)	Screen display conditions and presentation rate had no significant effect on math scores. Verbal scores were a quarter of a standard deviation higher with the larger, highest resolution display.
Bridgeman (2004)	Extra time added about 7 points to verbal scores and 7 points to quantitative scores. The accommodation appeared to have a greater impact on the quantitative scores of lower ability examinees.
Buehler (2002)	Results indicated that students with learning disabilities did not use significantly more time on the CAT/5, even when given the option. Students with disabilities did not receive any differential benefit from the use of the additional time accommodation. Although there were no differences in the reliability of the CAT/5 due to the accommodation of additional time, the validity of the CAT/5 was lower for students with learning disabilities who received additional time. The CTOPP was not found to be a useful predictor of students that would benefit from additional time on the CAT/5.

Burch (2002)	In comparison to students without LD, students with both reading and math disabilities experienced large accommodation boosts in the following conditions: CRT, V, and CA. Students with only reading disabilities did not receive an accommodation boost larger than students without LD under any condition.
Cahalan (2002)	In general, the revised SAT was to be positively correlated with FGPA for students who took the test with extended time accommodations for a learning disability. SAT scores were fairly accurate predictors of FGPA for students with learning disabilities. In the majority of cases when HSGPA was used along with SAT test scores, the predictive validity of FGPA was increased.
Choi (2002)	Item difficulty estimates did not appear to be the same across modes, particularly on the reading test and at the third grade level. When comparing identical items that were administered across both modes, computer items tended to have higher item difficulty estimates. Scrolling reading passages on computer screens seemed to have interfered with a student's test-taking behavior, particularly for younger students.
Cisar (2004)	Special educators and administrators tended to score higher than general and elective area teachers in their ability to distinguish and use assessment modifications.
Crawford (2004)	A significant interaction was found at grade 5 between length of time allotted for the assessment and the students' education classification. Grade 5 students performed significantly better on the 3-day writing assessment, with students in special education benefiting the most. The eighth-graders performed no better on the 3-day assessment than in the 30-minute assessment. Significant differences were reported across certain writing traits.
Dempsey (2003)	The verbal comprehension index was identified as the score that most closely predicts LSAT performance. This study found that scores earned under accommodated conditions are better than those earned under standard conditions.
Elbaum (2004)	As a group, students' test performance did not differ in the two conditions, and students with learning disabilities did not benefit more from the accommodation than students without learning disabilities. However, students with learning disabilities showed greater variability in their response to the accommodation.
Elliott (2003)	The performance of students with disabilities was highly similar to the performance of students without disabilities under standard time and extended time testing conditions. Overall, the provision of the accommodation, extended time, did not significantly improve scores of students with disabilities on the math test.
Elliott (2004)	The scores achieved in the extended time condition were higher than the scores achieved in the standard condition for all groups. However, the scores of students with disabilities did not improve significantly more than those of the students without disabilities when given extra time. A large proportion of survey respondents across all three groups expressed approval of the extended time condition.

Gagnon (2004)	No significant differences existed between teacher and principal reports of school-level curricular, assessment, and accountability policies. However, several statistically significant differences existed in school policies for schools that served students from a single district and those that served students from across a single state or more than one state. Approximately two-thirds of all the schools administered district and state assessments and most schools used their state's accommodation guidelines.
Hall (2002)	The study found that nearly 75% of fifth grade students with disabilities who participated in the MSPAP 2000 received test accommodations. Nearly half of these students received reading accommodations that invalidated the construct of the reading test, and almost a third of these students received writing accommodations that invalidated the language usage test. These reading and writing accommodations resulted in the reading and language usage scores of thousands of students with disabilities not being reported. Seventy-five percent of students with disabilities received accommodations and a third of these students met the satisfactory standard in the subject areas assessed. Also, although 25% of students with disabilities did not receive accommodations, about one-third of these students met the satisfactory standard.
Helwig (2002)	Elementary students with disabilities tended to perform better under the read aloud condition; elementary general education students did not appear to receive a similar benefit from the accommodation. For middle school students, no significant interactions were found.
Helwig (2003)	The teachers in the study were not effective in their recommendations of which students would, and would not, benefit from having math tests items read aloud. Teachers' ratings of their student's needs for testing accommodations coincided with actual students performance only half the time. The study found no connection between performance on reading and basic math skills tests and the need for accommodations.
Huynh (2002)	Accommodations provided on the separate form did not appear to substantially change the internal test structure. Students with disabilities taking the regular test form did not perform as well as other groups; students with disabilities taking the accommodated form performed as well as students without disabilities taking the regular form.
Huynh (2004)	It was found that the test structure remained rather stable across the three groups. Controlling for student background variables, disabled students under oral administration performed better than disabled students on the non-accommodated format. On the non-accommodated format, students with disabilities fared worse than general education students.
Idstein (2003)	Qualitative results show the better students do well in less time than it takes weaker students to achieve lower grades. Weaker students rely excessively on their dictionaries and do not trust themselves. Dictionary use does not effect the scores or test time of the better students, and may actually slow down and negatively effect the scores of weaker students.
Jackson (2003)	Scores did not differ among individual students due to demographic factors.



Janson (2002)	Students who received special education services and received accommodations experienced significant gains in scores in science and social studies in the year they were initially granted accommodations. There were substantial gains in science and social studies in 2000 for students initially receiving accommodations. There were significant gains in social studies and math scores in 2001 for students initially receiving accommodations.
Kappel (2002)	In general, no increase in scores was found when testing accommodations were used, and no differential response to using accommodations was found among subgroups or those without disabilities.
Katzman (2004)	The students with disabilities reported that they did not feel that they were prepared to take the MCAS because they believed that they were not taught the content on the test. Many students were not enrolled in courses that prepared them for the MCAS.
Kettler (2003)	Among fourth grade students, accommodations provided a larger effect for students with disabilities than students without disabilities on both the mathematics and the reading tests. Among eighth grade students, the effects of testing accommodations depended on the test content (math versus reading). The effects of testing accommodations on the math tests were somewhat higher for students with disabilities than for students without disabilities. Conversely, the effects of testing accommodations on the reading tests were slightly lower for students with disabilities than for students without disabilities.
Kobrin (2003)	The results suggest that computerized and paper-and-pencil reading comprehension tests may be more cognitively similar than originally thought. The only significant difference between computerized and paper-and-pencil tests was in the frequency of identifying important information in the passage.
Landau (2003)	Students performed better on five of the eight items when using the Talking Tactile Tablet, and performed the same on the remaining three. Using the Talking Tactile Tablet also yielded item difficulties that more closely resembled the item difficulties obtained by general education students during testing.
Macarthur (2004)	The results indicate that two-thirds (68%) of the students achieved 85% accuracy and more than one-third (40%) achieved 90% accuracy using dictation to a scribe or speech recognition software. Only 3 students (10%) were below 80% accuracy. Results for adults have been reported between 90% and 98%. Results also demonstrate that both dictation conditions helped students with learning disabilities produce better essays. Students with learning disabilities produced higher quality essays when using a scribe, then when using speech recognition software. Both adapted conditions were better in quality than handwritten essays.
McKevitt (2003)	The use of the read-aloud accommodation did not significantly improve the test performance of either group of students. Teachers as a group had neutral attitudes about testing and testing accommodations.
Meloy (2002)	Analyses revealed that students in both groups (LD-R and non-LD) achieved significantly higher test scores with the read aloud test administration.
Nickerson (2004)	The students felt that a majority of the accommodations used were helpful.

Pomplun (2002)	Analyses indicated that both forms of the computerized versions produced higher vocabulary scores than the paper-and-pencil format and one-form also had higher comprehension and total scores on the computerized version. These differences appeared to be related to the differences in response speed associated with use of a mouse to record responses as opposed to a pencil and answer sheet. Scores on the paper-and-pencil version and the computerized version had similar predictive power for course-placement.
Reed (2002)	Instances of learned helplessness and low motivation, a problem for LD and AD/HD students, were observed. To aid these students, test makers must do the following: Be cautious with the context, Ask the question clearly, and Repeat key words when possible in the response options.
Scheuneman (2002)	Almost 95% of students brought calculators to the November administration of the examination in both years. About 65% used their calculators on one third or more of the items. Group differences in the use of calculators were detected with girls using calculators more frequently than boys and Whites and Asian Americans using them more often than other racial groups. Although calculator presence, frequency of use, and calculator type were all correlated with test scores, this relation appears to be the result of the more able students using calculators differently from the less able students. Regression analyses revealed that a small percentage of the variance in test scores was accounted for by calculator access and type of calculator. Differential item functioning analyses (DIF) showed items favoring both frequent use and little use of calculators. Data concerning the rate of completion provided evidence that those using calculators less often were more likely to complete the exam.
Shriner (2003)	In this intervention study, training was found to increase the quality and extent of participation and accommodation documentation on the IEP. Correlations between what was documented on the IEP and what happened on the day of testing were highly variable. Although students' IEPs appeared to reflect individualized decisions, political and logistical factors limited the utility of the IEP and interfered with its actual implementation.
Tavani (2004)	Findings demonstrated non-significant performance score increases when students with learning disabilities who used accommodations were compared to those students who did not use accommodations.
Thornton (2002)	Overall, results suggest that LSAT scores earned under the nonstandard time condition are not comparable to LSAT scores earned under standard timing conditions. Results for individual subgroups were consistent with the overall group result.
Tindal (2002)	A main effect for both student classification and test administration was found for the elementary school students: Low achieving students outperformed students with IEPs and both groups benefited with a video-taped administration. For middle school students, a main effect was found for student classification; however, no main effects for the type of test administration (video versus standard).

Trammell (2003)	The impact of special accommodations on the subgroups revealed a significant improvement in grades for students with ADD and students with Learning Disabilities and ADD, but a drop for students with Learning Disabilities. Students with ADD and ADD and Learning Disabilities experienced an increase in grades with all types of accommodations conversely, students with learning disabilities experienced a drop in grades with each accommodation.
Weston (2003)	The findings revealed a statistical difference between the tests, and also between the two groups of students. Students with learning disabilities who are poor readers gained the most from the read-aloud accommodation. Results also suggest that the results on the accommodated test better match the teacher's estimations of the student's mathematical abilities.
Woods (2004)	The investigation found a low level of candidate need for a reader with candidate reading age and self-prediction being unreliable indicators of this need.

## Appendix G—Summary of Limitations Cited by Researchers

Author	Small Sample Size/Sample Too Narrow in Scope
Barton (2003)	This study included mostly mild to moderately disabled students. A replication with participating students with more severe disabilities, particularly severe physical disabilities, would certainly be beneficial.
Bridgeman (2003)	Although this study provides evidence that issues of screen size and resolution cannot be ignored, even larger studies are needed to understand fully the separate roles of screen size and resolution.
Burch (2002)	The study is limited by the small sample sizes of the groups. So, it is possible the students were not representative of the entire population of fourth-graders because the sample sizes were small.
Crawford (2004)	Further research using larger samples is needed on the effects of extended time for students with learning disabilities in the upper grades.
Hall (2002)	The study focuses only on fifth grade students, and the results may not generalize to students with disabilities in other grades or dissimilar disabilities, socio-economic statuses, etc.
Helwig (2002)	There were too few fifth-grade low-skill readers taking Form A in accommodated format to do meaningful analyses.
Idstein (2003)	The sample size was smaller than expected due to unpredictable attendance rates.
Jackson (2003)	Visual impairment is a low incidence disability and the number of possible participants is restricted. Many of the potential participants were eliminated because they were given the alternate assessment due to additional disabilities that affect student performance.
Janson (2002)	The study was conducted in a small school system in Tennessee. The study was limited to 448 students. Due to the small sample size of students who took the tests with accommodations it would be problematic to generalize the findings to a larger population.
Kappel (2002)	The sample size of 11 to 12 per group was possibly insufficient to detect effects that were present.
Kettler (2003)	One potential limitation of this study is that we examined only two test content areas and two grade levels. Specifically, we examined only mathematics and reading, although students are tested in science and social studies as well.
Landau (2003)	The small sample limited the analysis of the impact of the test accommodations on the psychometric properties of items.
McKevitt (2003)	This study focused solely on reading. Although the question of interest focused on a read-aloud accommodation, it was addressed only in the context of a reading test.
Pomplun (2002)	The ability to draw generalizations from these results could be limited by having students from only seven schools participate in the study.
Tavani (2004)	This study was intentionally limited with respect to its population and dependent variable. The target population was restricted to students in the 4 <sup>th</sup> , 8 <sup>th</sup> , and 12 <sup>th</sup> grades in the United States in 2000; consequently, this study may have sacrificed on its level of external validity.

<b>Author</b>	<b>Conflicting Results</b>
Tindal (2004)	Much of this research is tentative with conflicting overall test results: some findings show positive effects for all students, other findings reflect interactions between an accommodation and a population.
Weston (2003)	For non-disabled students the evidence is mixed and may be flawed by methodological problems. First, very low readers in the regular classroom did not seem to profit from the accommodation. Second, item content did not seem to affect general education students. Third, these students performed better on a number of items in the paper and pencil format.
<b>Author</b>	<b>Nonstandard Administration Across Proctors and Schools</b>
Huynh (2002)	South Carolina High School Exit Examination tests are un-timed; hence all students are permitted to take as long as they need to complete the tests regardless of whether they are disabled or not. Therefore, the findings of this study are not applicable to test administration modes that involve extended-time accommodations.
Huynh (2004)	Different school authorities made IEP and 504 accommodation decisions across grades 8 and 10; therefore, it is conceivable that a subset of this population should have been tested under oral administration at grade 8.
<b>Author</b>	<b>Confounding Factors</b>
Barton (2002)	Some students may be more accustomed to receiving the oral accommodation in their daily instruction and may therefore be practiced in test taking forms or environments that involve a good deal of listening.
Barton (2003)	A confound in this study was that students were not randomly assigned to discourse type nor to specific prompts within discourse type.
Cahalan (2002)	Some of these variations may be due to different populations of students used in the second sample. In sample two, colleges and universities were permitted to omit students for any reason including receiving services for a learning disability and having a FGPA less than 1.0.
Elbaum (2004)	The confounding of the accommodation with concomitant factors such as self-pacing and individual administration was a serious limitation.
Elliott (2004)	The students had more than enough time in the standard time condition likely diminished the impact of the accommodation of extra time. Also, nearly all students with disabilities receive multiple accommodations on district and statewide tests, thus the extended time accommodation when provided in isolation is contrived and not realistic.
Gagnon (2004)	Two limitations exist with the current study: (a) low response rate; and (b) differences in the characteristics of respondents versus nonrespondents.
Kobrin (2003)	An important limitation of this study is the lack of a time limit imposed on participants, because actual testing situations include time limits.
Macarthur (2004)	It is important to keep in mind that this study did not include extensive training in the use of speech recognition. Students received approximately 6 hours of individual instruction on training the software to recognize their speech and using it to compose essays.
Meloy (2002)	The read-aloud administration did not permit student self-pacing, and this procedure could have had an impact on students; maintaining attention to the test. Moreover, the administration was done by reading scripts for the various testings, whereas using a prerecorded tape could have reduced possible varying reader emphases.

Nickerson (2004)	Variables other than language proficiency and accommodations may affect performance on the SAT-9 (e.g. program model, access to curriculum, instruction, test preparation, etc.).
Reed (2002)	The use of the think-aloud procedure in this research may have in itself been beneficial for student performance. The process of self-explanation has been shown to improve students' problem-solving performance.
Trammell (2003)	Students with learning disabilities or learning disabilities plus ADD were not well matched with the accommodations they selected and were granted. This pitfall was the first and foremost limitation addressed in the design of the experiment and requires much further refinement and investigation.
<b>Author</b>	<b>Flaw in Research Design</b>
Bolt (2004)	This was not an experimental study, and there are subsequently limits to the inferences that can be made.
Cisar (2004)	The questionnaire that was developed for this study was not tested for construct or content validity. The validity of the test will indicate if it measured what it presumed to measure thus making the instrument more meaningful.
Elliott (2003)	When given twice as much time to work on the test, students neither took advantage of the extra time, nor showed significant gains in their scores. The students had more than enough time to complete the test diminishing the impact of the extra time accommodation.
Tindal (2002)	The sampling plan of the study was neither random nor stratified for teachers or students. Rather, teachers had been nominated for participation based on personal contacts of state department personnel through their own networks with principals and others in the local educational agencies.
<b>Author</b>	<b>No Limitations Mentioned</b>
Bielinski (2003)	
Bridgeman (2004)	
Choi (2002)	
Dempsey (2003)	
Hansen (2002)	
Helwig (2003)	
Katzman (2004)	
Schenueman (2002)	
Shriner (2003)	
Thornton (2002)	
Woods (2004)	
<b>Author</b>	<b>Not Applicable/Meta-analysis</b>
Sireci (2003)	
Tindal (2004)	



## Appendix H—Summary of Suggestions for Future Research (as recommended by authors)

Author	Investigate characteristics of accommodations themselves in further detail.
Barton (2003)	It is important to continue the research in the area of comparability of assessments that are administered online or on paper, both on scoring comparability and on the performance comparability of all students.
Bielinski (2003)	The findings in this study point to the need for further dialogue and more research on test accommodations. The opinions of those who influence policy and who are familiar with test accommodations vary too much to ignore.
Elbaum (2004)	The effects of different components of the accommodation need to be assessed separately.
Elliott (2004)	Researchers need extra time to answer this question: It is difficult to determine what a lack of boost in scores on accommodated test conveys about the effectiveness of the accommodation—specifically, did the accommodation provide access to the test so that the student’s true ability was assessed, or did the accommodation itself negatively affect the students’ performance?
Janson (2002)	Further research studies should be conducted in other Tennessee school systems to determine if accommodations, as provided in Tennessee, “level the playing field.”
Kappel (2002)	This study might also be expanded to include other factors, other accommodations. Also the effects of the Extended Time and Read-Aloud accommodation when administered in a group setting should also be investigated.
Kobrin (2003)	Future research should highlight the ways in which computerized tests may completely replace paper-and-pencil tests.
McKevitt (2003)	Future research examining the differential impact that decoding, fluency, or comprehension difficulties may have on reading test performance and the effects of accommodations would be useful.
Meloy (2002)	Future research on the read aloud accommodation is needed. Additional aspects of the read aloud procedures should be studied, and further refinements in design and sampling would be helpful.
Nickerson (2004)	An investigation into content validity, consequential validity and possible test bias is warranted for English language learners being assessed for academic achievement in the content areas of Reading/Language Arts and Mathematics with the SAT-9 tests.
Tavani (2004)	An investigation of a similar model utilizing NAEP databases is warranted in order to examine the impact these characteristics have on differing subject performances.
Weston (2003)	For any policy decision that contemplates providing students with accommodations, more research should be done to learn if the patterns shown in this study can be reproduced.



<b>Author</b>	<b>Investigate student factors contributing to accommodations use.</b>
Bridgeman (2003)	More research is needed to investigate the effects of a high stakes test on the psychological state of computer scrolling test takers. This study did not address that issue.
Crawford (2004)	Follow-up research studies investigating student's use of time during writing assessments will provide researchers with information related to differences across grades and educational classifications. The stakes associated with large-scale testing are too high to ignore the need for empirical evidence supporting the validity of multiple-day writing assessments.
Idstein (2003)	An observation based on interview data and subsequently discussed with classroom teachers, points to a possible correlation between general personality traits and dictionary use. This topic may warrant further investigation.
Katzman (2004)	Future research should examine the amount of support required to help students stay motivated.
Scheuneman (2002)	Differences in students' approach to problems when using calculators is an area where further investigation would be required.
Trammell (2003)	Students with learning disabilities in addition to ADD exhibited erratic decision making regarding accommodation requests. They may have been ill matched to the accommodation used during test taking. This pitfall requires much further refinement and investigation.
<b>Author</b>	<b>Improved Study Design or Study Replication</b>
Cahalan (2002)	More research is needed to investigate which factors contribute to the varied correlations in this study.
Helwig (2002)	The poor performance of some of the participants was, in fact, due to the distraction of the video when it was not needed, a logical solution would be an on-demand delivery system. A computer, audiotape, or live reading of only items selected by students on an individual basis would likely solve this problem. Further research in this area is warranted.
Landau (2003)	It is strongly suggested that in future studies participants be allowed to work with the Talking Tactile Tablet prior to testing and that more thorough beta-testing should be performed prior to testing.
Tindal (2002)	Further research may begin to utilize a videotaped administration along with other, more powerful and individualized accommodations that deal with setting and time.
Tindal (2004)	An effort needs to be made to improve consistency and systematicity in both practice and research, while maintaining clarity.
<b>Author</b>	<b>Study policy and accommodations hypotheses.</b>
Bolt (2004)	Research should continue to investigate accommodation decision-making and administration practices as well as Universal Design.
Gagnon (2004)	The results of this study indicate a need for more research concerning how the policies related to increased accountability are being implemented in special schools.
Reed (2002)	Additional clarification is needed for discussing the role of item or test intent in determining the appropriateness of an accommodation. Clear definitions of the constructs a given test measures should inform the practice of granting accommodations.

Sireci (2003)	Due to a wide variety of results stemming from experiential research it is suggested that a revision of the interaction hypotheses be proposed and that directions for future research and for improved test development and administration practices be proposed.
<b>Author</b>	<b>Replicate study with larger sample.</b>
Burch (2002)	Future studies should address the limitations of this study as well as the other studies in this area. The small sample sizes in this study create a need for replication.
Jackson (2003)	This study should be replicated to validate the results, with a larger sample across a variety of states, classifying adaptations into components identified in this study, and with more comparison to assessment results of non-disabled students.
Pomplun (2002)	Because the present study was based on students from only seven institutions, research studies should be continued, especially to support predictive validity of the computerized versions.
<b>Author</b>	<b>Investigate teacher factors related to accommodations selections.</b>
Cisar (2004)	Replication of this study in other states may be beneficial to determine if teachers can distinguish and know when to use adaptations nationwide.
Kettler (2003)	More research is needed on the apparently highly individualized nature of the impact of testing accommodations. What factors influence educators' selection of testing accommodations for specific students?
<b>Author</b>	<b>Investigate possible instructional uses for accommodations.</b>
Barton (2002)	Future analyses that seek to guide instructional effectiveness of students will augment this research. It would be interesting to look at each item that loaded on particular factors to see what qualitative characteristic they possess. If commonalities exist across items, such information may supplement instructional level information and types of approaches teachers may take with students.
<b>Author</b>	<b>Investigate Accommodation Practicality</b>
Macarthur (2004)	Future research should investigate the practical issues involved in using speech recognition in school settings and the impact of use over an extended time.
<b>Author</b>	<b>No Suggestions for Future Research Mentioned</b>
Barton (2003)	
Bridgeman (2004)	
Choi (2002)	
Dempsey (2003)	
Elliott (2003)	
Hall (2002)	
Hansen (2002)	
Hansen (2002)	
Helwig (2003)	
Huynh (2002)	
Huynh (2004)	
Shriner (2003)	

Thornton (2002)	
Woods (2004)	