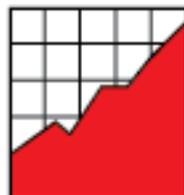


A Summary of the Research on the Effects of Test Accommodations: 2005-2006



**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 47

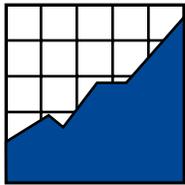
A Summary of the Research on the Effects of Test Accommodations: 2005-2006

April L. Zenisky
Stephen G. Sireci
Center for Educational Assessment
University of Massachusetts Amherst

August 2007

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

Zenisky, A. L., & Sireci, S. G. (2007). *A summary of the research on the effects of test accommodations: 2005-2006* (Technical Report 47). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.



**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. 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 | Sheryl S. Lazarus |
| Jason Altman | Kristi K. Liu |
| Manuel T. Barrera | Ross E. Moen |
| Laurene Christensen | Michael L. Moore |
| Marjorie Cuthbert | Rachel F. Quenemoen |
| Christopher J. Johnstone | Dorene L. Scott |
| Jane L. Krentz | Martha L. Thurlow, Director |

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

Six years have elapsed since the passage of the No Child Left Behind Act of 2001 (Public Law 107-110), and among its effects—principally on state accountability measures but also across other testing contexts from college admissions and professional credentialing to diagnostic/intelligence assessment, classroom evaluation, and beyond—is an increasing convergence of longtime policy and psychometric discussions about the use of various test accommodations and score interpretations from accommodated and non-accommodated administrations. At the same time, much work remains. The purpose of this report is to provide an update on the state of the research on testing accommodations as well as to identify promising areas of research to further clarify and enhance understanding of current and emerging issues. In 2005 and 2006, 32 published research studies on the topic of testing accommodations were found. Among the main points:

Purpose: The majority of the research included in this review sought to evaluate the comparability of test scores when assessments were administered with and without accommodations. The second most common purpose for research was to report on current accommodations practices (both in general and for populations exhibiting specific disabilities).

Types of assessments, content areas: Math and reading were the most common content areas included in the 2005-2006 research, and a wide variety of assessment types were used in these studies. Among academic measures, state criterion-referenced tests were common, as were miscellaneous intelligence and cognitive measures. Some studies also involved instruments developed for research purposes using publicly released items from various large-scale assessments such as the National Assessment of Educational Progress (NAEP), the Programme for International Student Assessment (PISA), and state tests.

Participants: Studies ranged from fewer than ten participants to several that involved tens of thousands of students, and spanned a range of grade levels from K-12 to college/university students, as well as one study that involved adult education.

Disabilities and accommodations: Learning disabilities were the most common disabilities exhibited by participants in the considered research, accounting for nearly half of the studies. Extended time (alone and bundled with other accommodations) was the single most studied accommodation, but oral accommodations (such as read-aloud and audiocassette presentation) were also considered in multiple studies, as was computerized administration.

Research design: Over 70% of the studies reported primary data collection on the part of the researchers, rather than drawing on existing archival data sets. Almost half of the

studies involved experimental or quasi-experimental designs. Researchers also drew on survey techniques and carried out literature meta-analyses.

Findings: Most of the oral presentation and timing accommodations empirically tested were found to have positive effects on scores, although some studies reported no effects for these accommodations. Among studies of the perception of different accommodations, researchers indicated that certain accommodations are more prevalent with some populations, that teacher training can affect accommodations practices in classrooms, and that what student Individualized Education Programs (IEPs) call for in terms of testing accommodations are not always the same as what ultimately is provided or what is used in instruction.

Limitations: Researchers often cited small sample size as well as a general lack of diversity as primary limitations of their research. Methodological issues relating to how accommodations were operationalized or experimentally implemented were also mentioned.

Directions for future research: A number of promising suggestions were noted, particularly with respect to varying or improving on research methods with respect to testing for the effects of specific accommodations and improving test development practices to reduce the need for accommodations. In many cases, researchers also found the results from their current studies raised many suggestions for further investigation, such as concurrent validity studies using other measures.

Our analysis across the studies identified a number of promising trends as well as opportunities for further advancing both research and practice. The focus across these studies on the use and effects of testing accommodations at different ages from elementary and secondary to post-secondary and adult education signals the importance of looking at differences in accommodations practices in different testing contexts, although increased diversity among research participants with respect to socioeconomic status or race/ethnicity is still needed.

Although many of the studies reported that accommodations use had some positive effect on test scores, variations across studies in the operational definitions of those accommodations does challenge the extent to which findings can be generalized across studies. Furthermore, even though much work is being done, another challenge for research is to construct true experiments to assess the effects of accommodations use on test scores and their consequences for students with and without disabilities alike.

Table of Contents

| | |
|----------------------------------------------------------------------------------------------------------------|----|
| Overview | 1 |
| Review Process | 2 |
| Results | 4 |
| Purposes of the Research | 4 |
| Research Type, Data Collection, and Research Designs..... | 5 |
| Assessment/Data Collection Focus | 7 |
| Content Area Assessed..... | 8 |
| Number of Research Participants (Total and Percent of Sample Consisting of Students with Disabilities | 8 |
| Grade Level..... | 10 |
| Disabilities Included in Research..... | 10 |
| Types of Accommodations in Reviewed Research | 11 |
| Research Findings..... | 11 |
| Limitations | 13 |
| Future Research | 13 |
| Discussion and Implications for Future Research | 14 |
| References..... | 19 |
| Appendix A: Research Purposes | 25 |
| Appendix B: Research Characteristics..... | 29 |
| Appendix C: Assessment/Instrument Characteristics | 31 |
| Appendix D: Participant and Sample Characteristics | 35 |
| Appendix E: Accommodations Studied | 37 |
| Appendix F: Research Findings..... | 41 |
| Appendix G: Limitations and Future Research | 45 |

Overview

Although the “standardized” in standardized testing may have multiple connotations, positive and negative alike, the term standardized is often described as a way to promote fairness in assessment by way of maintaining consistency in all aspects of test administration across test-takers. That said, according to the Common Core of Data from the National Center for Education Statistics, in the 2004-2005 school year (the most recent year for which these data are available) nearly 6 million of 48.7 million students in the United States had individualized education programs (IEPs) (National Center for Education Statistics, 2006). In many cases the disabilities that prompt these IEPs make it difficult for many students to perform to their full potential on tests under standard conditions, and so while not an exact barometer of test accommodation use, these statistics do indicate that on average across the states about 13-14% of elementary and secondary students have had teams of educators and specialists individually define their specific needs in instruction or assessment. One approach to assessment cannot always fit all because test-takers across many testing contexts often vary by more than just proficiency, due in part to the presence of one or more disabilities that can impact how they interact with and complete tasks in a testing situation. The use of test accommodations is often a necessity, as is the need for research-based policy to guide practice.

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999) define an accommodation as “an action taken in response to a determination that an individual’s disability requires a departure from an established testing protocol” (p. 101). More recently, researchers have referred to the accommodations as the means for eliminating construct-irrelevant variance, in other words, the variance associated with an extraneous feature of test administration (Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000). Others have concentrated on the notion that accommodations are test changes that maintain the validity of the scores that result from the testing process, by remaining true to the construct assessed. Numerous research approaches have been pursued to check that on the validity of scores produced under accommodated conditions (Thurlow, McGrew, Tindal, Thompson, Ysseldyke, Elliott, 2000; Sireci, Scarpati, & Li, 2005; Tindal, 1998), including single subject designs, “boost” studies, and “differential boost” studies.

Technical assistance providers and researchers have categorized and listed accommodations in several ways. For example, more than 70 accommodations in 8 categories (motivation, assistance prior to testing, scheduling, setting, directions, assistance during testing, use equipment/adaptive technology, and changes in format) were identified by Elliott, Kratochwill, and Schulte (1998) and placed into a checklist that they produced for IEP teams to use. Summaries of state policies show that there are probably hundreds of individual accommodations that can be identified, and that IEP teams have the option of identifying additional accommodations for individual

students, if needed (see, for example, Lazarus, Thurlow, Lail, Eisenbraun, & Kato, 2006). The specific accommodations that are used, how they are implemented, and the extent to which the scores from tests administered under standard and non-standard administrations are comparable are among the issues that are at the forefront of many conversations in many testing contexts today, including the states that must report on academic achievement for students with IEPs as part of No Child Left Behind (NCLB).

NCLB has placed a strong policy emphasis on students with disabilities by requiring that states focus on the performance of subgroups, both during their participation in state assessments and in national assessments. This focus is played out by requiring that the scores of subgroups be disaggregated and reported separately, as well as within the data reports of all other students, and that for accountability, they be treated in the same way—factored into accountability both separately and as part of the total group (and any other groups to which they belong). Beyond that, with new regulations (*Federal Register*, April 9, 2007), states must prepare accommodation guidelines that “identify the accommodations for each assessment that do not invalidate the score” as well as prepare IEP teams to “select, for each assessment, only those accommodations that do not invalidate the score” (Section 300.160(b)(2)). Within this context, the need for contributions to policy and psychometric understanding of the issues surrounding the use of test accommodations from researchers who are empirically studying these issues is at a critical point.

The purpose of this document is to provide a synthesis of the research on test accommodations published in 2005 and 2006. The research described here encompasses empirical studies of score comparability and validity studies as well as investigations into accommodations use and perceptions of their effectiveness. Taken together, the current research explores many of the issues surrounding test accommodations practices in both breadth and depth. Insofar as reporting on the findings of current research studies is a primary goal of this analysis, a second goal is to also identify areas requiring continued investigation in the future.

Review Process

To complete this review of the accommodations research published in 2005 and 2006, seven research databases were consulted, including Educational Resources Information Center (ERIC), PsychInfo, Academic Search Premier, Digital Dissertations, Education Complete, Expanded Academic ASAP, Educational Abstracts, and ISI Web of Science. In addition, two Web search engines were also used (Google and Google Scholar). Several other resources for research articles that were also searched for relevant publications were the archives of Behavioral Research and Teaching (BRT) at the University of Oregon (<http://brt.uoregon.edu/>), the Educational Policy Analysis Archives (EPAA; <http://epaa.asu.edu>), the National Center for Research on Evaluation, Standards, and Student Testing (CRESST; <http://www.cse.ucla.edu/>), the Wisconsin Center for

Educational Research (WCER; <http://www.wcer.wisc.edu/testacc>), and the Center for the Study of Assessment Validity and Evaluation (C-SAVE; <http://www.c-save.umd.edu/index.html>).

Finally, hand searches of relevant journals were conducted to ensure that no relevant articles were missed. Journals searched included: *Applied Measurement in Education*; *British Journal of Special Education*; *Educational and Psychological Measurement*; *Educational Measurement: Issues and Practice*; *Educational Psychologist*; *Educational Psychology*; *Educational Researcher*; *Exceptional Children*; *Journal of Educational Measurement*; *Journal of Learning Disabilities*; *Journal of Special Education*; *The Journal of Technology, Learning, and Assessment*; *Journal of Psychoeducational Assessment*; *Practical Assessment, Research, and Evaluation*; *Review of Educational Research*; and *School Psychology Review*. Presentations from professional conferences were not searched or included in this review, based on a preference to include only that research which (1) would be accessible to readers wanting to access the articles, and (2) had gone through the level of peer review typically required for publication in professional journals.

Within each of these research databases and publications archives, a sequence of search terms was used. Terms searched for this review were:

- accommodation(s)
- test *and* assess (*also* tests, testing, assessing, assessment) accommodation(s)
- test *and* assess (*also* tests, testing, assessing, assessment) changes
- test *and* assess (*also* tests, testing, assessing, assessment) modification(s)
- test *and* assess (*also* tests, testing, assessing, assessment) adaptation (adapt, adapting)
- student(s) with disability (disabilities) test *and* assess (*also* tests, testing, assessing, assessment)
- standards-based testing accommodations
- large-scale testing accommodations

The research documents from these searches were then considered for inclusion in this review with respect to several criteria. The decision was made to focus only on research published or defended in doctoral dissertations in 2005 and 2006. The scope of the research was limited to investigations of accommodations for regular assessment (hence, articles specific to alternate assessments, accommodations for instruction or learning, and universal design in general were not part of this review). In addition, research involving English language learners (ELLs) were only included if the focus was ELLs with disabilities.

Results

As a result of the search efforts, a total of 32 studies published between January 2005 and December 2006 met the criteria and are summarized in this review. Of these 32 studies, all but 6 appeared in refereed journals. Five of the six not from refereed journals were doctoral dissertations, and one was a published technical report. Seventeen of the studies involved an analysis of examinee responses to test questions in some way; nine used survey, interview, observation, or case study techniques to report on the use of test accommodations; and six involved reviewing literature and case law on testing accommodations or accommodations policies. A complete list of the research (researchers and full citations for each study included in this review) is given in the References.

Purposes of the Research

Several primary purposes were identified in the accommodations research published in 2005-2006 (see Table 1). Most commonly, these studies sought to investigate the effects of one or more test accommodations on students or items. This was the focus of over 40% of the studies. All but 4 of these 14 comparison studies involved students both with and without disabilities; 2 of the remaining studies looked at the results of assessments under standard and nonstandard administration conditions for students with disabilities only (Baker, 2006; Dolan, Hall, Bannerjee, Chun, & Strangman, 2005), and 2 varied test administration formats among students without disabilities (Higgins, Russell, & Hoffman, 2005; Horkay, Bennett, Allen, Kaplan, & Yan, 2006).

Table 1. Purposes of Reviewed Research

| Purpose | Number of Studies |
|----------------------------------------------------------------------------------|-------------------|
| Compare scores from standard/nonstandard administration conditions | 14 |
| Across students with and without disabilities (10 studies) | |
| Only students with disabilities (2 studies) | |
| Only students without disabilities (2 studies) | |
| Report on implementation practices and test accommodation use | 10 |
| Review test accommodation literature for effects on scores, assessment practices | 3 |
| Identify predictors of accommodation use | 3 |
| Study and/or compare perceptions of accommodation use | 2 |
| Total | 32 |

A full listing of the studies by purpose category including statements of purpose is provided in Appendix A.

The next most prevalent purpose in the reviewed research, involving 10 studies, was reporting survey, interview, or literature review results of accommodations use in different educational contexts, focusing specifically on implementation practices and institutional factors relating to accommodations use. Three of these studies were literature reviews of previous accommodations studies with respect to the effects of test accommodations on scores and assessment practices, and another three looked at ways to identify the need to use accommodations (Antalek, 2005; Gregg et al., 2005; Ofiesh, Mather, & Russell, 2005). Two articles (Lang et al., 2005; Packer, 2005) reported on perceptions of accommodations on the part of different stakeholder groups (parents, students, and educators in the former, and parents only in the latter).

Research Type, Data Collection, and Research Designs

There are several ways in which the research methods of these studies can be categorized. The first of these focuses on the status of each study as experimental, quasi-experimental, or non-experimental. A summary of studies by research type is given in Table 2, and detailed in Appendix B. In this categorization, an experiment (n=7) is characterized by random assignment of participants to at least one experimental condition. In contrast, the quasi-experiments (n=8) do not involve random assignment at all to any condition and instead are predicated on analyses of intact groups. Non-experimental studies (n=14) do not entail group comparisons or experimental manipulations of accommodations use.

Table 2. Research Type

| Research Type | Number of Studies |
|--------------------|-------------------|
| Experimental | 7 |
| Quasi-Experimental | 11 |
| Non-Experimental | 14 |

Research design was given additional scrutiny. For the studies involving group comparisons (the experimental and quasi-experimental studies) the research designs identified in Thurlow et al. (2000) were used to describe studies. These designs are described briefly here and are illustrated in Figure 1.

- *Design 1: Score comparability as a function of the presence/absence of a disability with equivalent test forms*
 Defining characteristics: equivalent forms, each participant completes all forms, random assignment to conditions within groups, includes students with and without disabilities.

- *Design 2: Score comparability as a function of the presence/absence of a disability with matched samples*
 Defining characteristics: single test form, each participant completes one form, matched samples, includes students with and without disabilities.
- *Design 3: Score comparability as a function of the use of an accommodation for a single disability*
 Defining characteristics: equivalent forms, each subject takes all forms, random assignment to conditions, includes only students with disabilities.
- *Design 4: Score comparability as a function of the use of an accommodation for subjects with disabilities*
 Defining characteristics: single test form, each participant completes one form, matched samples, includes only students with disabilities.

Figure 1. Research Designs 1, 2, 3, and 4 from Thurlow et al. (2000)

| | | Design 1 | | | | Design 2 | | | |
|---------------------|--|--------------------|---------|--------------------|---------|--------------------|---------|--------------------|---------|
| | | Disabilities | | No Disabilities | | Disabilities | | No Disabilities | |
| | | Group 1 | Group 2 |
| With Acc. | | Form A | Form B | Form A | Form B | Form A | | Form A | |
| Without Acc. | | Form B | Form A | Form B | Form A | | Form A | | Form A |
| | | Design 3 | | Design 4 | | | | | |
| | | Disability Group 1 | | Disability Group 2 | | Disability Group 1 | | Disability Group 2 | |
| With Acc. | | Form A | | Form B | | Form A | | | |
| Without Acc. | | Form B | | Form A | | | | Form B | |

Several other group designs for comparisons were also used in this research, and these were largely a variation on Design 2 (Bolt & Ysseldyke, 2006; Bruins, 2006; Huynh & Barton, 2006) and variations on Design 4 (Higgins et al., 2005; Horkay et al., 2006; Cohen, Gregg, & Deng, 2005). In addition, studies such as Gregg et al. (2005) and Shaftel, Belton-Kocher, Glasnapp, and Poggio (2006) administered the same tests to students with and without disabilities to identify predictors of accommodations needs.

Among the non-experimental studies, designs that were used included case studies (Horvath, Kampfer-Bohach, & Kearns, 2005; Rickey, 2005), literature reviews (Edgemon, Jablonski, & Lloyd, 2006; Meyen, Poggio, Seok, & Smith, 2006; Sahlen & Lehmann, 2006; Sireci, 2005; Sireci et al., 2005; and Stretch & Osborne, 2005), observations (Van Weelden & Whipple, 2005), and surveys (Cawthon, 2006; Cox, Herner, Demzyk, & Nieberding, 2006; Gibson, Haaeberli,

Glover, & Witter, 2005; Maccini & Gagnon, 2006; Packer, 2005).

A third and final characteristic of the techniques reported in accommodations research published in 2005-2006 is the source of the data, reflecting the decision of the researchers to use primary or archival/secondary data. In the former case, data collection is initiated and carried out by the researcher for the specific purpose of a study; the alternative is archival/secondary data, which is an available data set collected for a purpose other than research question. A cross-tabulation of data collection source level by research design is given in Table 3. A breakdown of research type, data collection, and research design information by reference is located in Appendix B.

Table 3. Studies by Research Designs and Data Collection Source

| | Research Design | Data Collection Source | | Total |
|----------------------------------------|--------------------------|------------------------|-----------|-----------|
| | | Primary | Archival | |
| Group comparison (15 studies total) | Design 1 | 5 | -- | 5 |
| | Design 2 | 2 | 3 | 5 |
| | Design 3 | 1 | -- | 1 |
| | Design 4 | 3 | 1 | 4 |
| | Other design | -- | 3 | 3 |
| Non-experimental (10 studies total) | Case study | 2 | -- | 2 |
| | Literature-based studies | -- | 6 | 6 |
| | Survey | 4 | 1 | 5 |
| | Observation | 1 | -- | 1 |
| Total | | 18 | 14 | 32 |

Assessment/Data Collection Focus

The accommodations research included here takes place in a wide variety of testing contexts, as indicated by the variety of instruments used in the studies (see Table 4). State criterion-referenced assessments, often used for NCLB purposes, were the most common data collection instruments involved in the studies (Bolt & Ysseldyke, 2006; Bruins, 2006; Cohen et al., 2005; Cox et al., 2006; Edgemon et al., 2006; Fletcher et al., 2006; Huynh & Barton, 2006; Meyen et al., 2006; and Shaftel et al., 2006). Researcher-developed survey instruments and interview protocols were the next most common data collection instruments used (Cawthon, 2006, Horvath et al., 2005; Lang et al. 2005; Maccini & Gagnon, 2006; Packer, 2005; Rickey, 2005; and Van Weelden & Whipple, 2005). Miscellaneous standardized academic achievement measures (a category that includes various Woodcock-Johnson subtests, Nelson-Denny Reading tests, and others) similarly accounted for over 20% of the studies reviewed (Antalek, 2005; Gregg et al., 2005; Lesaux et al., 2006; Ofiesh et al., 2005; Sahlen & Lehmann, 2006; Sireci et al., 2005; and Stretch & Osborne, 2005).

A number of other studies considered norm-referenced academic achievement tests such as the Stanford Achievement Test (SAT), ACT, and Graduate Record Examination (GRE) (Baker, 2006; Gibson et al., 2005; Kettler et al., 2005; Lang et al., 2005; Schnirman, 2005; and Sireci, 2005). Researcher-developed instruments were test forms created by the researchers for the express purpose of using them in their studies, most often using released test items from established testing programs such as the SAT, the National Assessment of Educational Progress (NAEP), and the Programme for International Reading and Language Arts Standards (PIRLS), and state assessments (Dolan et al., 2005; Higgins et al., 2005; Horkay et al., 2006; and Mandinach, Bridgeman, Cahalan-Laitusis, & Trapani, 2005). A listing of studies by assessment context of interest is given in Appendix C.

Table 4. Assessment/Data Collection Instruments

| Type | Number of Studies* |
|-----------------------------------------------------------------------|--------------------|
| State criterion-referenced assessment | 9 |
| Surveys/case study/interview protocols | 7 |
| Miscellaneous standardized academic achievement/intelligence measures | 7** |
| Norm-referenced academic achievement tests | 6*** |
| Researcher-developed academic measures | 4 |

* One study included more than one type of data collection method.

** Includes two literature reviews that were nonspecific about the tests used in the articles reviewed.

*** Includes one literature review that focused on accommodations use with tests for postsecondary admissions.

Content Area Assessed

Accommodations research published in 2005-2006 spanned a wide range of content areas. Mathematics and reading (along with assorted language arts constructs such as writing, spelling, and vocabulary, among others) were among the most often studied domains, as shown in Table 5. Other academic domains such as science, social studies, and music were also considered. Four studies of testing accommodations did not mention specific content areas. A complete list of content area or areas addressed in each study is provided in Appendix C.

Number of Research Participants (Total and Percent of Sample Consisting of Students with Disabilities)

A summary of the research participants is given in Table 6; this is further detailed for each study in Appendix D. Among the reviewed studies, the overall number of participants in the research varied from those that were small-scale studies, which included 10 or fewer individuals, to those that were very large-scale studies, which included over 300 individuals. The smallest study (Horvath et al., 2005) involved 9 research participants, while the largest reported data

Table 5. Academic Content Areas Involved

| Content Areas Assessed | Total* |
|--------------------------|--------|
| Mathematics | 17 |
| Reading | 14 |
| Misc. Language Arts** | 9 |
| Writing | 4 |
| Science | 1 |
| Social Studies | 1 |
| Civics/U.S. History | 1 |
| Music | 1 |
| No specific content area | 7 |

* Some studies included an examination of accommodations in more than one content area.

** Miscellaneous Language Arts assessment areas include *Language Usage*, *Verbal*, *Spelling*, *Listening*, and *Vocabulary*.

from over 107,000 examinees and six grade levels (Bolt & Ysseldyke, 2006). The proportion of participants in the research studies who were individuals with disabilities ranged from 0% (Higgins et al., 2005; Horkay et al., 2006) to 100% (Antalek, 2005; Baker, 2006; Dolan et al., 2005; Gibson et al., 2005; Horvath et al., 2005). Six studies reported data gathered from teachers, parents, schools, and states about individuals with disabilities and accommodations practices or use (Packer, 2005; Cawthon, 2006; Maccini & Gagnon, 2006; Rickey, 2005; Cox et al., 2006; Van Weelden & Whipple, 2005), while twenty addressed individual test-takers and five were literature reviews reporting on multiple studies with ranges of sample sizes and populations not individually reflected here. One involved legal cases.

Table 6. Cross tabulation of Sample Size by Percent of Individuals with Disabilities in Sample

| Total Number of Research Participants | Percent of Sample Consisting of Individuals with Disabilities | | | | | | N |
|---------------------------------------|---------------------------------------------------------------|--------|--------|---------|--------------|-----------------|----|
| | 0-24% | 25-49% | 50-74% | 75-100% | Not reported | Not applicable* | |
| 1-10 | -- | -- | -- | 2 | -- | 1 | 3 |
| 11-100 | -- | 1 | 2 | 1 | -- | 2 | 6 |
| 101-300 | 1 | 2 | 2 | 1 | -- | 2 | 8 |
| More than 300 | 3 | 1 | 2 | 1 | -- | -- | 7 |
| Not applicable* | -- | -- | -- | -- | 1 | 7 | 8 |
| N | 4 | 4 | 6 | 5 | 1 | 12 | 32 |

* These studies included (1) literature reviews of multiple studies where samples varied widely across the multiple studies included in each of the reviews, and (2) research studies that did not include students directly as the unit of analysis (e.g., they reported data from parents and/or teachers or aggregated results at the school or state level).

Grade Level

Most accommodations research that was completed involved K-12 students, with 13 studies involving elementary students, 15 focusing on middle school, and 15 also concerned with high school students (see Table 7). Specific grade levels for individual studies are reported in Appendix D, along with information on sample size and percent of sample with disabilities.

Table 7. Grade Level of Research Participants

| Education Level of Participants in Studies | Number of Studies * |
|---------------------------------------------------|----------------------------|
| Elementary School (K-5) | 13 |
| Middle School (6-8) | 15 |
| High School (9-12) | 15 |
| Postsecondary | 6 |
| Adults/Adult Education | 1 |
| Various, not specific | 2 |

* Counts include studies that spanned multiple grade levels.

Disabilities Included in Research

As shown in Table 8, learning disabilities were the most common disability for which testing accommodations were considered in the research, with 13 studies. Some of the lower incidence disabilities in the studies included deafblindness (Horvath et al., 2005) and different degrees of hearing loss (Cawthon, 2006), as well as Tourette's syndrome (Packer, 2005). Ten studies did not specifically identify the participants' disability, or included students without disabilities. A list of the disabilities reported by individual studies is included in Appendix D.

Table 8. Disabilities Reported in Research Participants

| Disabilities Observed in Research Participants | Number of Studies* |
|-----------------------------------------------------------------------------------------------------------------------------------|---------------------------|
| Learning disability | 13 |
| Disability not specified/general special needs students | 10 |
| Other disability (e.g., Physical/sensory disabilities, attention deficit disorder, health impairments, and multiple disabilities) | 8 |
| Emotional/Behavioral disability | 4 |
| Reading or Math deficit | 3 |
| Cognitive disability | 1 |

* Counts include studies involving students with multiple disabilities.

Types of Accommodations in Reviewed Research

Test accommodations experimentally or quasi-experimentally studied in the research fell into three categories: Presentation, Timing/Scheduling, and Setting. Response accommodations were not addressed in the research published in 2005-2006. Table 9 provides a brief summary of the accommodations studied in the research; this information is broken out by individual study in Appendix E. Extended time was the most frequently researched accommodation (Antalek, 2005; Baker, 2006; Bolt & Ysseldyke, 2006; Cohen et al., 2005; Lesaux et al., 2006; Mandinach et al., 2005; Ofiesh et al., 2005). Various implementations of oral administration including audiocassette presentation (Schnirman, 2005), read-aloud of proper nouns (Fletcher et al., 2006), and entire items (Bolt & Ysseldyke, 2006; Huynh & Barton, 2006), and computerized text-to-speech (Dolan et al., 2005) were examined in five studies. Two studies empirically studied the effects of accommodations as assigned by individual student IEPs (Bruins, 2006; Kettler et al., 2005), rather than focusing on specific individual accommodations.

Table 9. Accommodations in Reviewed Research

| Accommodation Category | Accommodation | Number of Studies |
|-------------------------------|---------------------------|--------------------------|
| Presentation | Oral administration | 5 |
| | Computer administration | 3 |
| | Scrolling vs. paging | 1 |
| Timing/Scheduling | Extended time | 7 |
| | Multiple day/sessions | 1 |
| | Separately timed sections | 1 |
| Setting | Small group/individual | 1 |
| As defined by students' IEPs | | 2 |
| Other | | 17* |

* The "Other" category is comprised of 17 studies where accommodations practices and use were explored but not experimentally (or quasi-experimentally) studied for their effects on test scores.

Research Findings

For those studies of the empirical effect of accommodations (see Table 10), none of the studies found any of the accommodations to have a negative impact on student scores, although for some accommodations the results were mixed. This was particularly the case for oral accommodations, computerized tests, and extended time. Overall, however, all of the timing accommodations reported a generally positive influence on scores. Specific study results by category are given in Appendix F.

Two studies focused on predicting the need for accommodations, and in both cases, the tests used were found to be helpful. The surveys of accommodations use indicated that for specific populations some accommodations are more prevalent and that teachers' use of accommodations

is often related to their training. From three studies, the selection and use of accommodations was found to be a complex undertaking requiring collaboration among stakeholders.

Table 10. Summary of Research Findings

| | Research Findings | Number of Studies* |
|-----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|---------------------------|
| Oral administration (read-aloud, audiocassette, text-to-speech) (n=5) | Positive effect on scores of students with disabilities when bundled with computer-based testing | 1 |
| | Positive effect on scores of students with disabilities when bundled with multiple sessions | 1 |
| | Associated with more DIF in Reading/Language Arts than Math | 1 |
| | No effect on scores | 2 |
| Computerized test (n=3) | Positive effect on scores of students with disabilities when bundled with oral administration | 1 |
| | No effect on scores | 2 |
| Scrolling vs. paging (n=1) | No effect on scores | 1 |
| Extended time (n=6) | Positive effect on scores of students with disabilities | 3 |
| | Positive effect on all student scores | 1 |
| | Extended time use did not explain observed Differential Item Functioning (DIF) | 1 |
| | DIF for read-aloud and extended time was consistent with DIF for read-aloud only | 1 |
| Multiple day/sessions (n=1) | Positive effect on scores of students with disabilities when bundled with oral administration | 1 |
| Separately timed sessions (n=1) | Positive effect on all student scores | 1 |
| Small group administration (n=1) | DIF for read-aloud and small group administration was consistent with DIF for read-aloud only | 1 |
| IEP-defined accommodations (n=3) | Positive effect on scores | 1 |
| | No positive effect | 1 |
| | Accommodations perceived as fair | 1 |
| Meta-analyses of Accommodated Conditions (n=3) | More empirical research needed | 3 |
| | Positive effect on scores of students with disabilities | 1 |
| Prediction of need for accommodations (n=2) | Tests were useful in prediction | 2 |
| Selection/implementation of accommodations (n=12) | Lack of alignment with IEP | 1 |
| | Some accommodations are more common than others | 4 |
| | Language characteristics have no disproportionate impact on students with disabilities | 1 |
| | Educators and institutions vary in accommodations use | 3 |
| | Determining appropriate assessment accommodations is a complex and collaborative undertaking | 3 |

* Some studies looked at more than one accommodation or reported more than one conclusion.

Limitations

Many of the studies included in this review noted at least one limitation to the research and findings. The limitations identified by the authors of the studies were classified as related to either the (1) research sample/participants (e.g., small sample size, lack of diversity), (2) test or testing context (e.g., number of items on the assessment instrument used), (3) methodology (e.g., decisions about study design, data collection, or data analysis), or (4) research results (e.g., unexpected findings that seem contradictory to established practice or other research). The numbers of studies in which each type of limitation was mentioned are summarized in Table 11; these are listed by study and category in Appendix G. As is evident in Table 11, the most frequently mentioned limitations focused on the samples used in the studies and methodology limitations.

Table 11. Limitations

| Limitation category | Number of Studies* |
|----------------------------|---------------------------|
| Sample characteristics | 16 |
| Methodology | 13 |
| Test/testing context | 8 |
| Results | 4 |
| No limitations listed | 11 |

*Many studies included more than one limitation.

Future Research

Future research directions identified in the accommodations studies published in 2005-2006 were categorized in terms of their recommendations for future studies to focus efforts on sample characteristics, tests and testing contexts, methodology, or results. A summary of future research by category is presented in Table 12; these suggestions are described more fully in Appendix G. Those suggestions categorized into the results category offered the most directions for future research, followed by those suggestions for improvements and advances in methodology.

Table 12. Future Research

| Future Research | Number of Studies* |
|-------------------------------------|---------------------------|
| Results | 19 |
| Methodology | 16 |
| Sample characteristics | 9 |
| Test/testing context | 7 |
| No future research directions given | 5 |

*Many studies listed more than one direction for future research.

Discussion and Implications for Future Research

The 32 studies included here present practitioners and researchers with a number of insights into both the current state of research on test accommodations and the directions that future research might take. At a broad level, most of the research published in 2005-2006 fell into one of two categories: (1) empirical studies of student scores from assessments administered under accommodated and non-accommodated conditions, and (2) research activities that were more descriptive in nature, aimed at identifying the accommodations used with different test populations or how accommodations use is perceived by different stakeholder groups.

Much of the research carried out to evaluate the comparability of scores from standard and nonstandard administrations included both students with and without disabilities (n=10), and implemented the full range of designs identified in Thurlow et al. (2000). Of the non-experimental work, most were surveys, but the research also included case studies and observations of assessment practices. Over 56% of the research studies (n=18) used primary data in their investigations rather than drawing on extant data sets.

As in previous summaries of accommodations research (Johnstone, Altman, Thurlow, & Thompson, 2006; Sireci et al., 2005), the domains of mathematics and language arts (specifically reading, but also writing and other related skills) were the most frequently studied content areas. Among the academic measures used in the studies, some were state tests used for NCLB purposes, but much research involved norm-referenced assessments, such as TerraNova (Gibson et al., 2005; Kettler et al., 2005; Lang et al., 2005) or the SAT.

The findings of the survey research studies presented in this review of 2005-2006 research reported that a wide variety of accommodations were in use for different student populations. It is interesting then, to note that there were just seven specific types of accommodations empirically studied and those were quite narrowly focused primarily in two categories (presentation and timing/scheduling). This finding was in contrast to earlier summaries of accommodations research by Johnstone et al. (2006) and Thompson, Blount, and Thurlow (2002), where there were 11 different accommodations within four categories reported as being studied empirically in each of those two reviews.

In the research summarized here the most common type of accommodation was timing/scheduling, with the specific accommodations studied including extended time, multiple testing sessions, and separately timed test sections. Presentation accommodations were the second most frequent type of accommodation provided. This category included computerized administration, oral administration (partial or whole read-aloud, computerized text-to-speech, and the use of audiocassettes), and scrolling or paging as the display method for passages. Five studies addressed specific accommodations in bundles (Fletcher et al., 2006; Dolan et al., 2005; Bolt &

Ysseldyke, 2006; Higgins et al., 2005; Mandinach et al., 2005), and only the design of Higgins et al. (2005) and Mandinach et al. (2005) permitted the results for the bundled accommodations to be discussed separately.

A wide range of disabilities and participant ages were reported in the participant samples in the accommodations research published in 2005-2006. Learning disabilities was the most common disability category included in the research, either singly (n=6) or in combination with other disabilities (n=7). About 30 percent of the studies did not report distinctions among the disabilities exhibited by students participating in the research. Other specific conditions that also emerged in the research included Tourette's syndrome, deafblindness, and deaf/hard-of-hearing. Research took place at all levels of education including postsecondary and adult schooling, and was evenly distributed across elementary, middle, and high school grade levels; indeed, about 80 percent of the research involved more than one grade level. Six studies were "very large" with participants numbering over 1,000 participants (and these analyses were carried out using extant testing program data); however, the majority of studies were moderate in scope, with data collected from 100 to 300 individuals.

Although this review of 2005-2006 accommodations research was not conducted as a formal meta-analysis, the patterns of research and results identified together raise a number of possible directions to inform future studies of accommodations use and the effects on student scores. These directions include (1) further study of extended time, (2) computers and assistive technology as accommodations, (3) the role of teachers, and (4) the interaction hypothesis.

The results for extended time, the most frequently researched accommodation in the 32 studies considered here, are generally consistent with the previous literature, where extended time had been shown to have a positive effect on the scores of students with disabilities. However, the emerging trend in elementary and secondary education toward the use of untimed tests for all students (as part of a larger strategy of integrating universal test design noted by Sireci et al., 2005), if it continues, may yet minimize the need for further study of the benefits of extended time test accommodations.

At the same time, while computerized administration is increasingly being considered for use across testing contexts, the research on different aspects of computer technology as test accommodations is not yet conclusive. This is due in part to operational challenges of implementing computer-based tests in practice or for research purposes. Nevertheless, computers do hold much promise for allowing students to use innovative formats and tailoring the presentation of the test to their individual needs (e.g., magnifying text, pacing in audio presentation). As reported in Johnstone et al. (2006), the computer as an accommodation investigated in the present research was not definitive. In addition, the presentation accommodation of scrolling or paging through passages did not have any effect on student scores one way or another, but further study com-

paring the effects for students with and without disabilities (rather than only students without disabilities) seems warranted. Ultimately, because of the range of ways that computerized tests can be formatted and administered for different purposes and content areas, a concerted program of research on operationally defining and evaluating computerized assessment accommodations, available on-demand, is needed. The review by Meyen et al. (2006) on the use of computerized-adaptive testing as a strategy for testing students with disabilities is likewise an important direction for future research, but computer use should be implemented carefully with respect to universal test design and with the goal of minimizing construct-irrelevant variance.

From the research involving teachers, significant variation among teachers was found in their familiarity with and use of different testing accommodations (Maccini & Gagnon, 2006). A disconnect was also found between the accommodations named in student IEPs, the accommodations used in everyday classroom instruction, and what was permissible for testing (Horvath et al., 2005). For student populations with specific disabilities, such as Tourette's syndrome (Packer, 2005) and deafness/hard of hearing (Cawthon, 2006), the research studies identified the most commonly used accommodations for those students.

The interaction hypothesis proposes that students with disabilities will benefit to a greater extent from accommodations than students without disabilities (i.e., there will be an interaction effect). This hypothesis was the topic of the article by Sireci et al. (2005), and the empirical results reported by Fletcher et al. (2006), Lesaux et al. (2006), and Kettler et al. (2005) provided support for the idea that students with disabilities needed accommodations and benefited from their use, while students without disabilities did not benefit from them (at least not to the same extent). In Fletcher et al. (2006), only students with disabilities benefited from the use of the orally-administered test given in multiple sessions, while Lesaux et al. (2006) and Kettler et al. (2005) found similar results for the extended time and IEP-assigned accommodations, respectively. In Sireci et al. (2005), evidence supporting a revision of the interaction hypothesis with respect to extended time was compiled. This revised hypothesis was based on the finding that both students with and without disabilities benefited from extended time, but the students with disabilities exhibited relatively greater score gains. This revision is consistent with differential boost theory (Fuchs & Fuchs, 2001; Thompson et al., 2002). Because accommodations represent departures from the standard testing protocol and almost always are considered to benefit only students with disabilities for whom they are appropriate, future research should continue to implement research designs that explicitly address the interaction hypothesis and differential boost to inform practice.

Although advancing understanding of the effects and use of testing accommodations, the authors of the 2005-2006 research on accommodations also took a critical eye to their own work and identified both limitations and findings deserving additional study. Many of the limitations they identified addressed aspects of research samples (small size, sample composition or homogene-

ity, lack of specific data, and motivation questions). Study design issues were also mentioned by several researchers including Dolan et al. (2005), who pointed out that the accommodations were tested in such a way that the interaction hypothesis was not evaluated. Both Huynh and Barton (2006) and Kettler et al. (2005) cited limitations related to the variations in how different accommodations can be operationalized and the extent to which such differences limit generalizability. One limitation across the studies of the effects of accommodations is the use of predominantly multiple-choice items in the measurement instruments. In fact, some studies, such as Cohen et al. (2005) eliminated constructed-response items to simplify the analyses. Given that Koretz and Hamilton (2000) found differences between the performance of students with disabilities' performance on multiple choice and constructed response items, future research should further evaluate potential differential impact of accommodations on these different item formats. While multiple choice items are certainly common in many assessments, other formats such as short-answer and extended-answer items are being used in state tests for K-12 students. In the future, studies of accommodations should look at strategies for implementing accommodations across more mixed-format tests.

The reviews of test accommodations issues completed by Sireci et al. (2005), Sireci (2005), and Stretch and Osborne (2005), respectively, were focused on the interaction hypothesis, score comparability and interpretation, and extended time accommodations, but together offered many important directions for future study. How accommodations are operationalized is one area where greater definition or clarification may be warranted, as is improved guidance for users of scores from accommodated and non-accommodated administrations about appropriate test score inferences.

Great diversity exists both with respect to the individuals requiring assessment accommodations and the range of accommodations available. The test accommodations research published in 2005-2006 and in previous years amply reflects that diversity, but such diversity does not easily lend itself to consensus on policy for valid testing practice. The completion of more well-constructed meta-analyses of specific accommodations is one strategy that researchers should consider, in addition to further empirical study of specific accommodations with different—both heterogeneous and homogeneous—student populations.

Bridging research and practice is ultimately no easy task, but at this point of reflection, taking stock of what has been learned from the 2005-2006 and previous years' studies is critical. The accommodations research findings to date offer advances in knowledge about the effects of accommodations, but in 2005-2006, as in previous years, variations across operational definitions, tests, populations, settings, and contexts still curb all but the most general policy implications. Decisions surrounding the use of testing accommodations involve increasingly high-stakes consequences, and yet interpreting scores from accommodated and non-accommodated administrations remains, in many cases, as much art as science. Johnstone et al. (2006) and others have

noted previously that broader changes and innovations in testing practices may help to lessen the need for accommodations for students with disabilities; this may be accomplished by revisiting the testing experience for all students, such as making tests untimed across the board. Still, additional, experimentally-designed research to identify best practices for operational testing and the communication of that information to interested researchers, educators, policymakers, parents, students with disabilities themselves, and other consumers, in clear and concise terms will help to ensure that students with and without disabilities alike are assessed equitably by methods that reflect the best that research and practice together can offer.

The assessment policies of NCLB strongly emphasize including all students in assessments and require disaggregated reporting for students with disabilities and other groups. These policies also emphasize obtaining valid measures of students' performance. For many students, valid measurement means providing accommodations that do not change the construct measured, but make the test more accessible to them. Thus, the need for understanding what the research on test accommodations tells us is more important than ever before. It will be essential to continue to review and summarize the research conducted in this area, and to question whether changes in assessment and accommodations policies need to be made. It may also be important to explore new designs and new hypotheses as research moves forward to address the policy implications of research findings in this area.

References

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

Antalek, E. E. (2005). The relationships between specific learning disability attributes and written language: A study of the performance of learning disabled high school subjects completing the TOWL-3. *Dissertation Abstracts International*, 65 (11), 4098 A. Retrieved August 5, 2006 from Digital Dissertations database.

Baker, J. S. (2006). Effect of extended time testing accommodations on grade point averages of college students with learning disabilities. *Dissertation Abstracts International*, 67 (1), 574 B. Retrieved August 5, 2006 from Digital Dissertations database.

Bolt, S. E., & Ysseldyke, J. E. (2006). Comparing DIF across math and reading/language arts tests for students receiving a read-aloud accommodation. *Applied Measurement in Education*, 19 (4), 329-355.

Bruins, S. K. (2006). Investigating how students with disabilities receiving special education services affect the school's ability to meet adequate yearly progress. *Dissertation Abstracts International*, 66 (8), 2889 A. Retrieved August 5, 2006 from Digital Dissertations database.

Cawthon, S. W. (2006, Summer). National survey of accommodations and alternate assessments for students who are deaf or hard of hearing in the United States. *Journal of Deaf Studies and Deaf Education*, 11, (3), 337-359.

Cohen, A. S., Gregg, N., & Deng, M. (2005). The role of extended time and item content on a high-stakes mathematics test. *Learning Disabilities Research & Practice*, 20 (4), 225-233.

Cox, M. L., Herner, J. G., Demczyk, M. J., & Nieberding, J. J. (2006). Provision of testing accommodations for students with disabilities on statewide assessments. *Remedial and Special Education*, 27 (6), 346-354.

Dolan, R. P., Hall, T. E., Bannerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test design: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *The Journal of Technology, Learning, and Assessment*, 3 (7). Retrieved August 5, 2006, from <http://escholarship.bc.edu/jtla/>.

Edgemon, E. A., Jablonski, B. R., & Lloyd, J. W. (2006). Large-scale assessments: A teacher's guide to making decisions about accommodations. *Teaching Exceptional Children*, 38 (3), 6-11.

- Elliott, S. N., Kratochwill, T. R., & Schulte, A. G. (1998). The assessment accommodation checklist: Who, what, where, when, why, and how? *Teaching Exceptional Children, 31* (2), 10-14.
- Fletcher, J. M., Francis, D. J., Boudousquie, A., Copeland, K., Young, V., Kalinowski, S., & Vaughn, S. (2006). Effects of accommodations on high-stakes testing for students with reading disabilities. *Exceptional Children, 72* (2), 136-150.
- Fuchs, L. S., & Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research & Practice, 16*, 174-181.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., & Karns, K. M. (2000). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review, 29*(1), 65-85.
- Gibson, D., Haaeberli, F. B., Glover, T. A., & Witter, E. A. (2005). Use of recommended and provided testing accommodations. *Assessment for Effective Intervention, 31* (1) [Special issue: Testing Accommodations: Research to Guide Practice], 19-36.
- Gregg, N., Hoy, C., Flaherty, D. A., Norris, P., Coleman, C., Davis, M., & Jordan, M. (2005). Decoding and spelling accommodations for postsecondary students with dyslexia—It's more than processing speed. *Learning Disabilities—A Contemporary Journal, 3* (2), 1-17.
- Higgins, J., Russell, M., & Hoffman, T. (2005). Examining the effect of computer-based passage presentation on reading test performance. *The Journal of Learning, Technology, and Assessment, 3* (4). Retrieved August 5, 2006, from <http://escholarship.bc.edu/jtla/>.
- Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment, 5* (2). Retrieved January 4, 2007 from <http://www.jtla.org>.
- Horvath, L. S., Kampfer-Bohach, S., & Kearns, J. F. (2005). The use of accommodations among students with deafblindness in large-scale assessment systems. *Journal of Disability Policy Studies, 16* (3), 177-187.
- Huynh, H., & Barton, K. E. (2006). Performance of students with disabilities under regular and oral administrations of a high-stakes reading examination. *Applied Measurement in Education, 19*(1), 21-39.
- Johnstone, C.J., Altman, J., Thurlow, M., & Thompson, S. J. (2006). *A summary of the research on the effects of tests accommodations: 2002 through 2004* (Technical Report 45). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

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

Koretz, D., & Hamilton, L. (2000). Assessment of students with disabilities in Kentucky: Inclusion, student performance, and validity. *Educational Evaluation and Policy Analysis*, 22(3), 255-272.

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

Lazarus, S.S., Thurlow, M.L., Lail, K.E., Eisenbraun, K.D., & Kato, K. (2006). *2005 state policies on assessment participation and accommodations for students with disabilities* (Synthesis Report 64). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Available at <http://education.umn.edu/nceo/OnlinePubs/Synthesis64/default.html>

Lesaux, N. K., Pearson, M. R., & Siegel, L. S. (2006). The effects of timed and untimed testing conditions on the reading comprehension performance of adults with reading disabilities. *Reading and Writing*, 19, 21-48.

Maccini, P., & Gagnon, J. C. (2006). Mathematics instructional practices and assessment accommodations by special and general educators. *Exceptional Children*, 72(2), 217-234.

Mandinach, E. B., Bridgeman, B., Cahalan-Laitusis, C., & Trapani, C. (2005). The impact of extended time on SAT test performance. *Research Report No 2005-8*. New York, NY: The College Board.

Meyen, E., Poggio, J., Seok, S., & Smith, S. (2006). Equity for students with high-incidence disabilities in statewide assessments: A technology-based solution. *Focus on Exceptional Children*, 38(7), 1-8.

National Center for Education Statistics. (2006). *Common Core of Data (CCD): School Years 2004 Through 2005*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Ofiesh, N., Mather, N., & Russell, A. (2005). Using speeded cognitive, reading, and academic measures to determine the need for extended test time among university students with learning disabilities. *Journal of Psychoeducational Assessment*, 23, 35-52.

Packer, L. E. (2005). Tic-related school problems: Impact on functioning, accommodations, and interventions. *Behavior Modification, 29*(6), 876-899.

Rickey, K. M. (2005). Assessment accommodations for students with disabilities: A description of the decision-making process, perspectives of those affected, and current practices. *Dissertation Abstracts International, 67*(1), 145 A. Retrieved August 5, 2006 from Digital Dissertations database.

Sahlen, C. A. H., & Lehmann, J. P. (2006). Requesting accommodations in higher education. *Teaching Exceptional Children, 38*(3), 28-34.

Schnirman, R. K. (2005). The effect of audiocassette presentation on the performance of students with and without learning disabilities on a group standardized math test. *Dissertation Abstracts International, 66*(6), 2172 A. Retrieved August 5, 2006 from Digital Dissertations database.

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment, 11*(2), 105-126.

Sireci, S. G. (2005). Unlabeling the disabled: A perspective on flagging scores from accommodated test administrations. *Educational Researcher, 34*(1), 3-12.

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

Stretch, L. S., & Osborne, J. W. (2005). Extended test time accommodations: Directions for future research and practice. *Practical Assessment, Research, and Evaluation, 10*(8). Retrieved August 5, 2006 from <http://pareonline.net/pdf/v10n8.pdf>.

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. Available at <http://education.umn.edu/NCEO/OnlinePubs/Technical34.htm>.

Thurlow, M. L., McGrew, K.S., Tindal, G., Thompson, S. L., Ysseldyke, J. E., & Elliott, J. L. (2000). *Assessment accommodations research: Considerations for design and analysis* (Technical Report 26). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Available at <http://education.umn.edu/NCEO/OnlinePubs/Technical26.htm>.

Tindal, G. (1998). *Models for understanding task comparability in accommodated testing*. A publication for the Council of Chief State School Officers, Washington, DC. Retrieved May 19, 2006, from the World Wide Web: <http://education.umn.edu/nceo/OnlinePubs/Acomm/Task-Comparability.htm>

VanWeelden, K., & Whipple, J. (2005). Preservice teachers' predictions, perceptions, and actual assessment students with special needs in secondary general music. *The Journal of Music Therapy*, 42(3), 200-221.

Appendix A

Research Purposes

Table A-1. Purpose Category: Compare Scores from Standard/Nonstandard Administration Conditions for Students With and Without Disabilities

| Author(s) | Stated Research Purpose |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Bolt & Ysseldyke (2006) | Examine the extent to which read-aloud accommodation allows for better measurement on a math test than a reading test. |
| Bruins (2006) | Determine (1) if there was a significant difference in the performance of general education students and special education students on the test, (2) if testing accommodations equal the testing performance of students with disabilities when scores are compared to nondisabled peers, and (3) the impact of including students with disabilities as a separate subgroup when calculating adequate yearly progress. |
| Cohen et al. (2005) | Investigate the influence of extended time and content knowledge on the performance of individuals taking a statewide math test with and without accommodations. |
| Fletcher et al. (2006) | Address interaction hypothesis by evaluating accommodations specifically designed to minimize the impact of word recognition difficulties on a high-stakes reading comprehension test, comparing the performance of students with word decoding problems with the performance of students with average word decoding ability. |
| Huynh & Barton (2006) | Examine the effect of oral administration accommodations on test structure and student performance on a reading test. |
| Kettler et al. (2005) | Examine the effects of IEP-assigned testing accommodations on mathematics and reading test scores. |
| Lesaux et al. (2006) | Examine the effects of extra time on reading comprehension performance of individuals with reading disabilities. |
| Mandinach et al. (2005) | Explore the impact of providing standard time, time-and-a-half with and without section breaks, and double time without specified section breaks on verbal and math SAT. |
| Schnirman (2005) | Conduct an empirical investigation of the effects of audiocassette presentation by comparing the performance of students with LD and students from general education, as well as establish the relationship, if any, between the level of knowledge of mathematics vocabulary and the benefit of audiocassette presentation for students with LD. |
| Shaftel et al. (2006) | Evaluate the impact of language characteristics in mathematics test items on student performance for students with disabilities and ELLs as well as general education students. |

Table A-2. Purpose Category: Compare Scores from Standard/Nonstandard Administration Conditions for Students with Disabilities

| Author(s) | Stated Research Purpose |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Baker (2006) | Investigate the relationship between the use of extended time testing accommodations and academic achievement in students with learning disabilities. |
| Dolan et al. (2005) | Investigate the potential of computer-based read-aloud testing accommodations, focusing on computer-based testing with text-to-speech as an approach for providing individualized support to students with learning disabilities during multiple-choice testing. |

Table A-3. Purpose Category: Compare Scores from Standard/Nonstandard Administration Conditions for Students Without Disabilities

| Author(s) | Stated Research Purpose |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Higgins et al. (2005) | Examine differences in performance when two different computer-based test formats and a traditional paper-and-pencil based format are used to present reading passages. |
| Horkay et al. (2006) | Investigate the comparability of scores for paper and computer versions of an eighth-grade writing test. |

Table A-4. Purpose Category: Report on Implementation Practices and Test Accommodations Use

| Author(s) | Stated Research Purpose |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Cawthon (2006) | Report the results from the National Survey of Accommodations and Alternate Assessments for Deaf and Hard-of-Hearing Students in the United States. |
| Cox et al. (2006) | Discuss accommodations-related research findings from a three-year federally funded study, examining accommodations policies and discipline rates in all fifty states. |
| Edgemon et al. (2006) | Provide recommendations and guidelines for accommodations decision-making, in addition to offering a framework for special educators to use in selecting accommodations that permit students with disabilities to demonstrate knowledge, competence, and learning on large-scale assessments. |
| Gibson et al. (2005) | Explore factors that potentially influence the implementation of recommended testing accommodations, with respect to (1) accommodations recommended through the IEP process, (2) accommodations recommended by the teacher, and (3) accommodations provided in the testing sessions. |
| Horvath et al. (2005) | Describe the use of accommodations among students with deafblindness both in general curriculum and during statewide assessments. |
| Maccini & Gagnon (2006) | Answer questions about what specific instructional practices do special and general education teachers reportedly use for students with learning disabilities (LD) and emotional or behavioral disabilities (EBD) during both instruction on and when assessing basic math computation skills and problem-solving tasks, and what factors predict the number of instructional practices and assessment accommodations general and special education teachers reportedly make for students with LD and EBD. |
| Meyen et al. (2006) | Explain a technology-based option (adaptive testing) that allows for the construction of tests tailored to the knowledge and skill attributes of individual examinees. |
| Rickey (2005) | Examine the implementation of the requirements of the 1997 IDEA Amendments mandating inclusion of students with disabilities, with the use of appropriate accommodations, in state and district assessments. |
| Sahlen & Lehman (2006) | Identify the considerations that students and postsecondary institutions address during legal cases involving accommodations requests. |
| VanWeelden & Whipple (2005) | Examine preservice teachers' predictions and perceptions of students with special needs' level of mastery of specific music education concepts and actual grades achieved by these students using alternate assessments and testing accommodations. |

Table A-5. Purpose Category: Review Literature on Test Accommodations for Effects on Scores and Assessment Practices

| Author(s) | Stated Research Purpose |
|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sireci (2005) | Review the psychometric issues regarding flagging test scores taken under non-standard conditions, discuss accommodations research in college admissions testing, and provide suggestions for determining when scores should be flagged. |
| Sireci et al. (2005) | Review numerous studies that focused on the effects of accommodations on test performance to see if students with disabilities benefited from accommodations relative to their nondisabled peers. |
| Stretch & Osborne (2005) | Summarize and discuss current research on extended time testing, particularly with respect to implications for assessment. |

Table A-6. Purpose Category: Identify Predictors of the Need for Test Accommodation(s)

| Author(s) | Stated Research Purpose |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Antalek (2005) | Determine if visual-motor processing speed is the most effective predictor of the need for extended time on complex writing tasks, or if other learning disability attributes could have a similar or more significant relationship upon the successful completion of a written task within a specific time allotment. |
| Gregg et al. (2005) | Examine the relationship between specific Woodcock-Johnson III Cognitive and Achievement clusters across populations with and without dyslexia, identify the strongest WJ II cognitive and linguistic predictors for decoding, spelling, and reading fluency, across samples with and without dyslexia, and discuss the implications of the findings for assessment and accommodations practices for secondary and postsecondary students. |
| Ofiesh et al. (2005) | Examine the relationship between scores on speeded cognitive and academic tests and the need for the accommodation of extended test time for normally achieving students and students with learning disabilities. |

Table A-7. Purpose Category: Study or Compare Perceptions of Accommodation Use

| Author(s) | Stated Research Purpose |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Lang et al. (2005) | Examine student, parent, and teacher perceptions of the use of testing accommodations and the relationship between student perceptions of testing accommodations and their disability status and grade level. |
| Packer (2005) | Provide data on (1) parental perceptions of how children's tics might impair specific academic activities and determine the impact of tic improvement on academic functions, (2) parental impressions on improvement of peer relationships if tics improved or remitted, and (3) how school personnel attempted to respond to tic-related problems and to determine the perceived effectiveness of these strategies. |

Appendix B

Research Characteristics

Table B-1. Research Types, Designs, and Data Sources

| Research Type | Studies | Group Design | | | | | Non-Exp. Design | Data Source |
|----------------------------|----------------------------|--------------|---|---|---|--------------|-----------------|-------------|
| | | 1 | 2 | 3 | 4 | Other Design | | |
| Experiment (n=7) | Kettler et al. (2005) | ✓ | | | | | | Primary |
| | Ofiesh et al. (2005) | ✓ | | | | | | Primary |
| | Schnirman (2005) | ✓ | | | | | | Primary |
| | Mandinach et al. (2005) | | ✓ | | | | | Primary |
| | Fletcher et al. (2006) | | ✓ | | | | | Primary |
| | Dolan et al. (2005) | | | | ✓ | | | Primary |
| | Higgins et al. (2005)* | | | | | ✓ | | Primary |
| Quasi-Experiment (n=11) | Lang et al. (2005) | ✓ | | | | | | Primary |
| | Lesaux et al. (2006) | ✓ | | | | | | Primary |
| | Bolt & Ysseldyke (2006)** | | ✓ | | | | | Archival |
| | Bruins (2006)** | | ✓ | | | | | Archival |
| | Huynh & Barton (2006)** | | ✓ | | | | | Archival |
| | Antalek (2005) | | | | ✓ | | | Primary |
| | Baker (2006) | | | | ✓ | | | Archival |
| | Horkay et al. (2006)*** | | | | ✓ | | | Primary |
| | Cohen et al. (2005)**** | | | | | ✓ | | Archival |
| | Gregg et al. (2005)***** | | | | | ✓ | | Archival |
| | Shaftel et al. (2006)***** | | | | | ✓ | | Archival |
| Non-Experiment (n=14) | Cawthon (2006) | | | | | | Survey | Primary |
| | Cox et al. (2006) | | | | | | Survey | Archival |
| | Gibson et al. (2005) | | | | | | Survey | Primary |
| | Maccini & Gagnon (2006) | | | | | | Survey | Primary |
| | Packer (2005) | | | | | | Survey | Primary |
| | VanWeelden&Whipple (2005) | | | | | | Observation | Primary |
| | Edgemon et al. (2006) | | | | | | Lit. review | Archival |
| | Meyen et al. (2006) | | | | | | Lit. review | Archival |
| | Sahlen & Lehmann (2006) | | | | | | Lit. review | Archival |
| | Sireci (2005) | | | | | | Lit. review | Archival |
| | Sireci et al. (2005) | | | | | | Lit. review | Archival |
| | Stretch & Osborne (2005) | | | | | | Lit. review | Archival |
| | Horvath et al. (2005) | | | | | | Case study | Primary |
| | Rickey (2005) | | | | | | Case study | Primary |

* Design 4 except all participants were students without disabilities.

** Design 2 with only one group without disabilities (no accommodations).

***Design 4 except all participants were students without disabilities and two groups received accommodations.

**** Design 4 except students with disabilities took accommodated test; students without disabilities took nonaccommodated.

***** Both students with disabilities and students without disabilities took the same tests to identify predictors of accommodations need.

Appendix C

Assessment/Instrument Characteristics

Table C-1. Assessment/Instrument Types and Specific Assessments/Instruments Used

| Studies | Researcher-developed survey/interview protocols | Miscellaneous academic achievement/intelligence measures | Norm-referenced academic achievement tests | State criterion-referenced assessment | Researcher-developed tests |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------------------|-----------------------------------------------------|
| Antalek (2005) | | <i>Test of Written Language (3rd Ed.)</i> | | | |
| Baker (2006) | | | SAT | | |
| Bolt & Ysseldyke (2006) | | | | Unspecified state's large-scale assessment | |
| Bruins (2006) | | | | <i>Idaho Standards Achievement Test</i> | |
| Cawthon (2006) | <i>National Survey of Accommodations and Alternate Assessments for Students who are Deaf or Hard of Hearing in the United States</i> | | | | |
| Cohen et al. (2005) | | | | <i>Florida Comprehensive Assessment Test</i> | |
| Cox et al. (2006) | | | | Various state NCLB assessments | |
| Dolan et al. (2005) | | | | | Released NAEP items |
| Edgemon et al. (2006) | | | | Various state NCLB assessments | |
| Fletcher et al. (2006) | | | | <i>Texas Assessment of Knowledge and Skills (practice form)</i> | |
| Gibson et al. (2005) | | | <i>TerraNova</i> | | |
| Gregg et al. (2005) | | <i>Woodcock-Johnson III (Various)</i> | | | |
| Higgins et al. (2005) | | | | | Released NAEP, PIRLS, and NH state assessment items |

| Studies | Researcher-developed survey/interview protocols | Miscellaneous academic achievement/intelligence measures | Norm-referenced academic achievement tests | State criterion-referenced assessment | Researcher-developed tests |
|-------------------------|---------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|----------------------------------------------------|-----------------------------------|
| Horkay et al. (2006) | | | | | NAEP items |
| Horvath et al. (2005) | Student, parent, and teacher interviews; student observations | | | | |
| Huynh & Barton (2006) | | | | <i>South Carolina High School Exit Examination</i> | |
| Kettler et al. (2005) | | | <i>TerraNova</i> (research forms) | | |
| Lang et al. (2005) | Student, parent, and teacher surveys | | <i>TerraNova</i> (research forms) | | |
| Lesaux et al. (2006) | | <i>Woodcock-Johnson, Wide Range Achievement Test, Wechsler Adult Intelligence Scale (Various)</i> | | | |
| Maccini & Gagnon (2006) | Teacher survey of assessment accommodations | | | | |
| Mandinach et al. (2005) | | | | | Released SAT items |
| Meyen et al. (2006) | | | | Various state NCLB assessments | |
| Ofiesh et al. (2005) | | <i>Kaufman Brief Intelligence Test, Weschler Adult Intelligence Scale, Woodcock-Johnson, Nelson Denny (Various)</i> | | | |
| Packer (2005) | Parental survey of school experiences | | | | |
| Rickey (2005) | Student, parent, and teacher interviews about accommodation practices/use | | | | |
| Sahlen & Lehmann (2006) | | Various college course assessments | | | |

| Studies | Researcher-developed survey/interview protocols | Miscellaneous academic achievement/intelligence measures | Norm-referenced academic achievement tests | State criterion-referenced assessment | Researcher-developed tests |
|-----------------------------|--------------------------------------------------------|-----------------------------------------------------------------|---------------------------------------------------|----------------------------------------------|-----------------------------------|
| Schnirman (2005) | | | <i>Iowa Tests of Basic Skills</i> | | |
| Shaftel et al. (2006) | | | | <i>Kansas General Assessments</i> | |
| Sireci (2005) | | | <i>SAT, GRE, ACT</i> | | |
| Sireci et al. (2005) | | Various | | | |
| Stretch & Osborne (2005) | | Various | | | |
| VanWeelden & Whipple (2005) | Pre-service teachers survey of accommodations use | | | | |
| Total | 7 | 7 | 6 | 9 | 4 |

Table C-2. Content Areas Assessed

| Author(s) | Math | Reading | Writing | Other LA* | Science | Social Studies | Civics/ US History | Music | Not Specific | N |
|---------------------------|------|---------|---------|-----------|---------|----------------|--------------------|-------|--------------|---|
| Antalek (2005) | | | ✓ | | | | | | | 1 |
| Baker (2006) | ✓ | | | ✓ | | | | | | 2 |
| Bolt & Ysseldyke (2006) | ✓ | ✓ | | ✓ | | | | | | 3 |
| Bruins (2006) | ✓ | ✓ | | ✓ | | | | | | 3 |
| Cawthon (2006) | ✓ | ✓ | | | | | | | | 2 |
| Cohen et al. (2005) | ✓ | | | | | | | | | 1 |
| Cox et al. (2006) | ✓ | ✓ | | | | | | | | 2 |
| Dolan et al. (2005) | | | | | | | ✓ | | | 1 |
| Edgemon et al. (2006) | | | | | | | | | ✓ | - |
| Fletcher et al. (2006) | | ✓ | | | | | | | | 1 |
| Gibson et al. (2005) | ✓ | ✓ | | | | | | | | 2 |
| Gregg et al. (2005) | | ✓ | | ✓ | | | | | | 2 |
| Higgins et al. (2005) | | ✓ | | | | | | | | 1 |
| Horkay et al. (2006) | | | ✓ | | | | | | | 1 |
| Horvath et al. (2005) | | | | | | | | | ✓ | - |
| Huynh & Barton (2006) | | ✓ | | | | | | | | 1 |
| Kettler et al. (2005) | ✓ | ✓ | | | | | | | | 2 |
| Lang et al. (2005) | ✓ | ✓ | | | | | | | | 2 |
| Lesaux et al. (2006) | ✓ | ✓ | | ✓ | | | | | | 3 |
| Maccini & Gagnon (2006) | ✓ | | | | | | | | | 1 |
| Mandinach et al. (2005) | ✓ | | | ✓ | | | | | | 2 |
| Meyen et al. (2006) | | | | | | | | | ✓ | - |
| Ofiesh et al. (2005) | ✓ | ✓ | ✓ | ✓ | | | | | | 4 |
| Packer (2005) | | | | | | | | | ✓ | - |
| Rickey (2005) | | | | | | | | | ✓ | - |
| Sahlen & Lehmann (2006) | | | | | | | | | ✓ | - |
| Schnirman (2005) | ✓ | | | | | | | | | 1 |
| Shaffel et al. (2006) | ✓ | | | | | | | | | 1 |
| Sireci (2005) | ✓ | | | ✓ | | | | | | 2 |
| Sireci et al. (2005) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | | 6 |
| Stretch & Osborne (2005) | | | | | | | | | ✓ | - |
| VanWeelden&Whipple (2005) | | | | | | | | ✓ | | 1 |
| N | 17 | 14 | 4 | 9 | 1 | 1 | 1 | 1 | 7 | |

* Other Language Arts assessment areas include Language Usage, Verbal, Spelling, Listening, and Vocabulary.

Appendix D

Participant and Sample Characteristics

Table D-1. Unit of Analysis, Total Sample Sizes (Students, Parents, Schools, Articles, and Teachers), Grade/Education Level, and Types of Disabilities

| Unit of Analysis | Studies (Year) | Sample Size | Percent of Sample with Disabilities | Grade/Education Level | Types of Disabilities Exhibited * |
|------------------|-------------------------|------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|------------------------------|---------------------------------------|
| Students | Antalek (2005) | 67 | 100% | High School | LD |
| Students | Baker (2006) | 127 | 100% | College (1 st yr) | LD |
| Students | Bolt & Ysseldyke (2006) | 16,447 gr.3 16,634 gr.4 16,849 gr.7 15,108 gr.8 13,672 gr.10 12,299 gr.11 | 70% gr. 3 70% gr. 4 70% gr. 7 67% gr. 8 63% gr. 10 59% gr. 11 | 3, 4, 7, 8, 10, 11 | LD, PD, OD |
| Students | Bruins (2006) | 70 gr. 4 82 gr. 8 88 gr. 10 | 50% | 4, 8, 10 | Type not documented |
| Students | Cohen et al. (2005) | 2,500 | 50% | 9 | LD |
| Students | Dolan et al. (2005) | 10 | 100% | 11, 12 | LD |
| Students | Fletcher et al. (2006) | 182 | 50% | 3 | RD (dyslexia) |
| Students | Gibson et al. (2005) | 354 | 100% | 4, 8 | LD, CD, EBD, PD (visual), OD (autism) |
| Students | Gregg et al. (2005) | 201 | 50% | College | RD (dyslexia) |
| Students | Higgins et al. (2005) | 219 | 0% | 4 | No disabilities |
| Students | Horkay et al. (2006) | 4,133 | 0% | 8 | No disabilities |
| Students | Horvath et al. (2005) | 9 | 100% | 4, 7, 8, 9 | PD (deafblindness) |
| Students | Huynh & Barton (2006) | 89,319 | 4% | 10 | PD, EBD, LD |
| Students | Kettler et al. (2005) | 118 gr. 4 78 gr. 8 197 total | 42% gr. 4 50% gr. 8 | 4, 8 | Type not documented |
| Students | Lang et al. (2005) | 152 gr. 4 142 gr. 8 294 total | 42% gr. 4 43% gr. 8 | 4, 8 | Type not documented |
| Students | Lesaux et al. (2006) | 64 | 34% | Adults | RD |
| Students | Mandinach et al. (2005) | 1,929 | 14% | 11 | LD, OD (ADHD) |
| Students | Ofiesh et al. (2005) | 84 | 51% | College | LD |
| Students | Schnirman (2005) | 48 | 50% | Middle School | LD |
| Students | Shaftel et al. (2006) | ~2,000 gr.4 ~2,000 gr.7 ~2,000 gr.10 | ~30-40% per grade | 4, 7, 10 | LD, OD |
| Parents | Packer (2005) | 69 | Not applicable | (Children aged 6 -17) | PD (tics) |

| Unit of Analysis | Studies (Year) | Sample Size | Percent of Sample with Disabilities | Grade/ Education Level | Types of Disabilities Exhibited * |
|-------------------------|---------------------------|-------------------------|--------------------------------------------|-------------------------------------------------------|------------------------------------------|
| Schools | Cawthon (2006) | 264 | Not applicable | (Children ranged in grade from 1st-12 th) | PD (deaf/hard of hearing) |
| Articles | Edgemon et al. (2006) | Not applicable | Not applicable | Elementary, Middle, High School | Type not documented |
| Articles | Meyen et al. (2006) | Not applicable | Not applicable | College | Type not documented |
| Articles | Sireci (2005) | 10 | Not applicable | High School, College | Type not documented |
| Articles | Sireci et al. (2005) | 59 | Not applicable | Not applicable | Type not documented |
| Articles | Stretch & Osborne (2005) | 42 | Not applicable | Not applicable | Nonspecific, LD |
| Teachers | Maccini & Gagnon (2006) | 179 | Not applicable | High School | LD, E/BD |
| Teachers/ IEP teams | Rickey (2005) | 9 | Not applicable | Middle School | Type not documented |
| Teachers | VanWeelden&Whipple (2005) | 15 | Not applicable | Middle School | E/BD |
| Legal Cases | Sahlen & Lehmann (2006) | 8 | Not applicable | College | Type not documented |
| States | Cox et al. (2006) | 18 ES 17 MS 16 HS | Not reported | Elementary, Middle, High School | Type not documented |

* Key:

LD (Learning Disability)

PD (Physical Disability)

RD (Reading Deficit)

CD (Cognitive Disability)

EBD (Emotional or Behavioral Disability)

OD (Other Disability)

Appendix E

Accommodations Studied

Table E-1. Accommodations Researched by Study

| Studies (Year) | Experimental Accommodations | | | | | | | | Other* |
|----------------------------|-----------------------------|----------|----------|---------------------|----------|----------|----------|----------|-----------|
| | Presentation | | | Timing / Scheduling | | | Setting | Other | |
| | Oral | CBT | S/P | Ext | MS | STS | SG | IEP | |
| Fletcher et al. (2006)** | ✓ | | | | ✓ | | | | |
| Huynh & Barton (2006) | ✓ | | | | | | | | |
| Schnirman (2005) | ✓ | | | | | | | | |
| Dolan et al. (2005)** | ✓ | ✓ | | | | | | | |
| Bolt & Ysseldyke (2006)** | ✓ | | | ✓ | | | ✓ | | |
| Higgins et al. (2005)*** | | ✓ | ✓ | | | | | | |
| Horkay et al. (2006) | | ✓ | | | | | | | |
| Antalek (2005) | | | | ✓ | | | | | |
| Baker (2006) | | | | ✓ | | | | | |
| Cohen et al. (2005) | | | | ✓ | | | | | |
| Lesaux et al. (2006) | | | | ✓ | | | | | |
| Mandinach et al. (2005)*** | | | | ✓ | | ✓ | | | |
| Ofiesh et al. (2005) | | | | ✓ | | | | | |
| Bruins (2006) | | | | | | | | ✓ | |
| Kettler et al. (2005) | | | | | | | | ✓ | |
| Cawthon (2006) | | | | | | | | | ✓ |
| Cox et al. (2006) | | | | | | | | | ✓ |
| Edgemon et al. (2006) | | | | | | | | | ✓ |
| Gibson et al. (2005) | | | | | | | | | ✓ |
| Gregg et al. (2005) | | | | | | | | | ✓ |
| Horvath et al. (2005) | | | | | | | | | ✓ |
| Lang et al. (2005) | | | | | | | | | ✓ |
| Maccini & Gagnon (2006) | | | | | | | | | ✓ |
| Meyen et al. (2006) | | | | | | | | | ✓ |
| Packer (2005) | | | | | | | | | ✓ |
| Rickey (2005) | | | | | | | | | ✓ |
| Sahlen & Lehmann (2006) | | | | | | | | | ✓ |
| Shaftel et al. (2006) | | | | | | | | | ✓ |
| Sireci (2005) | | | | | | | | | ✓ |
| Sireci et al. (2005) | | | | | | | | | ✓ |
| Stretch & Osborne (2005) | | | | | | | | | ✓ |
| VanWeelden&Whipple (2005) | | | | | | | | | ✓ |
| Total | 5 | 3 | 1 | 7 | 1 | 1 | 1 | 2 | 17 |

Oral = Oral Presentation (partial or whole)

CBT = Computer-Based Test

S/P = Scrolling or Paging on Computerized Test

ExT = Extended Time

MS = Multiple Sessions

SG = Small Group/Individual Administration

STS = Separately Timed Sections

IEP = Various accommodations were implemented as per individual student IEPs.

* The seventeen studies in the "Other" category include research activities where student performance or accommodations practices and use were explored but not experimentally (or quasi-experimentally) studied.

** These studies examined the effects of multiple accommodations in bundles.

*** These studies examined the effects of multiple accommodations separately.

Table E-2. Specifications for Table F-1: Nature of Accommodations Research by Study

| Studies (Year) | Specification |
|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Fletcher et al. (2006) | Students with and without dyslexia completed a test under accommodated (in combination: two sessions, oral reading of proper nouns, and oral reading of comprehension stems) or non-accommodated (single administration, no oral reading of proper nouns or comprehension stems) conditions. |
| Huynh & Barton (2006) | Students with disabilities completed a test under accommodated (oral administration) or non-accommodated (no oral administration) conditions, and students without disabilities completed a test under non-accommodated (no oral administration) conditions. |
| Schnirman (2005) | Students with and without learning disabilities completed equivalent forms of a test under accommodated (audiocassette read-aloud) and non-accommodated conditions. |
| Dolan et al. (2005) | Students with disabilities completed equivalent forms of a test under accommodated (in combination: computer-based administration with text-to-speech technology) and non-accommodated (paper) conditions. |
| Bolt & Ysseldyke (2006) | Students with disabilities completed a test under accommodated (read-aloud, with or without extended time and small group/individual administration) or non-accommodated (no read-aloud, no extended time or small group/individual administration) conditions, and students without disabilities completed a test under non-accommodated (no read-aloud) conditions. |
| Higgins et al. (2005) | Students without disabilities completed a test under accommodated (either on computer with scrolling through passages or on computer with paging through passages) or non-accommodated (paper) conditions. |
| Horkay et al. (2006) | Students without disabilities completed a test under accommodated (computer-based administration) and non-accommodated (paper) conditions. |
| Antalek (2005) | Students with and without learning disabilities were administered a test under non-accommodated (timed) conditions, but were given extra time if tasks were not completed in that time. |
| Baker (2006) | Students with disabilities' scores were compared on whether the individuals chose to complete classroom tests under accommodated (extra time) or non-accommodated (standard time) conditions. |
| Cohen et al. (2005) | Students with disabilities received extra time accommodations while students without disabilities completed the under standard conditions/no accommodations. |
| Lesaux et al. (2006) | Students with and without reading disabilities completed a battery of tests under accommodated (untimed) and non-accommodated (timed) conditions. |
| Mandinach et al. (2005) | Students with and without disabilities completed a multi-part test under accommodated (either) (1) 1 ½ time with separate timing for individual sections, (2) 1 ½ time with no separate timing for sections, or (3) double time) or non-accommodated (standard time) conditions. |
| Ofiesh et al. (2005) | Students with and without disabilities completed a battery of tests under accommodated (untimed) and non-accommodated (timed) conditions. |
| Bruins (2006) | Students with disabilities completed a test under accommodated (as assigned by their IEPs) or non-accommodated (standard) conditions, and students without disabilities completed a test under non-accommodated (standard) conditions. |
| Kettler et al. (2005) | Students with disabilities took a test under accommodated conditions (as assigned by their IEPs) and students without disabilities took a test under non-accommodated conditions. |
| Lang et al. (2005) | Students with and without disabilities were placed into matched pairs and administered tests under accommodated (as assigned by IEPs of the SwD) and non-accommodated conditions, and were asked to respond to survey questions about the experience. Teachers and parents were also surveyed. |
| Cawthon (2006) | Survey of schools and programs regarding processes for identifying assessment accommodations as well as implementation and use. |
| Cox et al. (2006) | State policies on accommodations and assessment participation rates for students with disabilities across states were examined. |
| Edgemon et al. (2006) | Guidelines for accommodations use in schools are provided for educators. |
| Gibson et al. (2005) | The AAC was used to create a common framework across districts to compare assessment accommodations recommended by IEPs, recommended by teachers, and actually provided during testing. |

| Studies (Year) | Specification |
|---------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Gregg et al. (2005) | Students with and without disabilities completed a battery of tests under standard conditions to try and identify predictors of the need for decoding and spelling accommodations. |
| Horvath et al. (2005) | Survey of students, parents, and teachers regarding processes for identifying assessment accommodations as well as implementation and use. |
| Maccini & Gagnon (2006) | Survey of teachers regarding processes for identifying assessment accommodations as well as implementation and use. |
| Meyen et al. (2006) | The use of computer adaptive testing as an assessment accommodation is suggested. |
| Packer (2005) | Survey of parents regarding processes for identifying assessment accommodations as well as implementation/use. |
| Rickey (2005) | Survey of students' IEP teams regarding processes for identifying assessment accommodations as well as implementation/use. |
| Sahlen & Lehmann (2006) | Eight court cases regarding requests for accommodations in higher education context are reviewed. |
| Shaftel et al. (2006) | Students with and without disabilities completed a test under non-accommodated conditions. |
| Sireci (2005) | 10 articles on the effects of flagging scores from accommodated administrations were analyzed. |
| Sireci et al. (2005) | 59 articles on the effects of various accommodations relative to the interaction hypothesis were analyzed. |
| Stretch & Osborne (2005) | 42 articles on the effects of extended time were analyzed. |
| VanWeelden&Whipple (2005) | Survey of preservice teachers' assessment practices. |

Appendix F

Research Findings

Table F-1. Findings for Oral Accommodations

| | |
|-------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Oral Accommodations had a positive effect on scores of students with disabilities when bundled with CBT</i> | |
| Dolan et al. (2005) | Scores on the computerized-oral test were significantly increased over paper scores when passages were longer than 100 words in length. |
| <i>Oral Accommodations had a positive effect on scores of students with disabilities when bundled with multiple sessions</i> | |
| Fletcher et al. (2006) | Only SwD benefited from the accommodations, showing a significant increase in average performance and a 7-fold increase in the odds of passing; results supported the interaction hypothesis. |
| <i>Oral Accommodations were associated with more DIF in Reading/Language Arts than Math</i> | |
| Bolt & Ysseldyke (2006) | A greater portion of DIF items were identified for those students receiving read-aloud accommodations on a reading/language arts test than a math test. Read-aloud accommodations were found to be associated with greater measurement incomparability for reading/language arts than math. |
| <i>Oral Accommodations had no effect on scores</i> | |
| Huynh & Barton (2006) | After controlling for major background variables, the performance of students with disabilities under oral administration conditions was comparable to that of students with disabilities who took the test under regular administration conditions. The internal structure of the HSEE test remained stable across students with disabilities and students without disabilities. |
| Schnirman (2005) | No statistically significant differences were found between performance of students with disabilities and students without disabilities. |

Table F-2. Findings for Computerized Test

| | |
|-------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| <i>Computerized Test had a positive effect on scores of students with disabilities when bundled with oral accommodations</i> | |
| Dolan et al. (2005) | Scores on the computerized-oral test were significantly increased over paper scores when passages were longer than 100 words in length. |
| <i>Computerized Test had no effect on scores</i> | |
| Higgins et al. (2005) | There were no significant difference in reading comprehension scores across testing modes. |
| Horkay et al. (2006) | Results showed no mean significant differences between paper and computer delivery. |

Table F-3. Findings for Scrolling vs. Paging

| | |
|------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| <i>Scrolling vs. Paging had no effect on scores</i> | |
| Higgins et al. (2005) | There were no significant difference in reading comprehension scores across testing modes. |

Table F-4. Findings for Extended Time

| <i>Extended Time had a positive effect on scores of students with disabilities</i> | |
|------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Antalek (2005) | The majority of the subjects took additional time and their scores on the task improved significantly, indicating a relationship between learning disabilities and the completion of academic tasks within an allotted time frame. |
| Baker (2006) | The group that used extended time accommodations had an average first year GPA that was 0.39 points higher (statistically significant) than the group that did not use accommodations. The use of extended time accounts for 11% of variance in full year GPA and 7% of overall GPA. |
| Lesaux et al. (2006) | Under timed conditions there were significant differences between performance of students with disabilities and students without disabilities. All of the students with disabilities benefited from extra time, but students without disabilities performed comparably under timed and untimed conditions. Also, students with disabilities (less severe) performed comparably to students without disabilities in untimed conditions. |
| <i>Extended Time had a positive effect on all student scores</i> | |
| Mandinach et al. (2005) | Results indicated that time and a half with separately timed sections benefits students with disabilities and students without disabilities, though some extra time improves performance and too much may be detrimental. Extended time benefits medium and high ability students but provides little or no advantage to low-ability students. |
| <i>Use of Extended Time did not explain DIF</i> | |
| Cohen et al. (2005) | Some items exhibited DIF under accommodated (extended time) conditions, but students for whom items functioned differently were not accurately characterized by their accommodation status but rather content knowledge. |
| <i>DIF for read-aloud and extended time was consistent with DIF for read-aloud only</i> | |
| Bolt & Ysseldyke (2006) | Read-aloud accommodations and extended time were found to be associated with a comparable level of DIF relative to the use of read-aloud only, and these results were consistent across both reading and math. |

Table F-5. Findings for Multiple Days / Sessions

| <i>Multiple Days/Sessions had a positive effect on scores of students with disabilities when bundled with oral admin.</i> | |
|----------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Fletcher et al. (2006) | Only students with disabilities benefited from the accommodations, showing a significant increase in average performance and a 7-fold increase in the odds of passing; results supported the interaction hypothesis. |

Table F-6. Findings for Separately Timed Sessions

| <i>Separately Timed Sessions had a positive effect on all student scores</i> | |
|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mandinach et al. (2005) | Results indicated that time-and-a-half with separately timed sections benefits students with disabilities and students without disabilities, though some extra time improves scores; too much may be detrimental. Extended time benefits medium/high ability students but provides little or no advantage to low-ability students. |

Table F-7. Findings for Small Group Administration

| | |
|-------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>DIF for read-aloud and small group administration was consistent with DIF for read-aloud only</i> | |
| Bolt & Ysseldyke (2006) | Read-aloud accommodations and small group administration were found to be associated with a comparable level of DIF relative to the use of read-aloud only, and these results were consistent across both reading and math. |

Table F-8. Findings for IEP-Assigned Accommodations

| | |
|----------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>IEP-Assigned Accommodations had a positive effect on scores</i> | |
| Kettler et al. (2005) | Students with disabilities benefited from accommodations more than students without disabilities, and the differential benefit was higher on Reading than Math. |
| <i>Effect of IEP-Assigned Accommodations had no positive effect</i> | |
| Bruins (2006) | Significant differences were found between the performance of general education students and students with disabilities, and the use of IEP-assigned accommodations did not have a positive effect on scores of students with disabilities. |
| <i>IEP-Assigned Accommodations are perceived as fair</i> | |
| Lang et al. (2005) | Parents and teachers perceive accommodations as fair and valid for students with disabilities. More students with disabilities than students without disabilities indicated that accommodations made test condition easier, more comfortable, and better indicator of knowledge. |

Table F-9. Findings for Meta-Analyses of Accommodations Practices

| | |
|---------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>More empirical research needed</i> | |
| Sireci (2005) | Current research and practice with respect to flagging scores from accommodated administrations is insufficient. |
| Sireci et al. (2005) | Research does not provide clear guidance because of the variety of accommodations studied, how they are operationalized in research, and variations in samples. |
| Stretch & Osborne (2005) | Recommendations for research: Find better estimates of ability; determine if tests are appropriate; consider inclusion of students with disabilities in samples; understand tentative nature of scores from accommodated tests; weigh quality of information source. |
| <i>Accommodations have a positive effect on scores of students with disabilities</i> | |
| Sireci et al. (2005) | Accommodation (extended time) tends to have a positive effect on scores of students with disabilities. Accommodation (oral) tends to have a positive effect on scores of students with disabilities . |

Table F-10. Findings for Prediction of Need for Accommodations

| | |
|----------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Tests of interest aid in identifying need for accommodations</i> | |
| Gregg et al. (2005) | Study provides strong evidence for the usefulness of the WJ III Cognitive Abilities clusters in predicting reading decoding and spelling performance of the postsecondary population with dyslexia. |
| Ofiesh et al. (2005) | The findings indicated significant group differences on all speeded cognitive, reading, and academic tests (with few exceptions). The WJ III Reading Fluency and Academic Fluency tests were the best predictors of students with disabilities needing extra time. |

Table F-11. Findings for Selection and Implementation of Accommodations

| | |
|------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Lack of alignment with IEP</i> | |
| Horvath et al. (2005) | Provided accommodations were not always tailored to their needs; class, test, and IEP accommodations did not always match up. |
| <i>Some accommodations are more common than others</i> | |
| Cawthon (2006) | The most prevalent test accommodations reported by schools/programs for students with disabilities (deaf/hard-of-hearing) included extended time, interpreter for directions, and a separate location. Read-aloud and signed Q-R were prevalent also but used more in math assessments than reading. Mainstream students with disabilities used accommodations more than those in schools for deaf/school or district programs. |
| Cox et al. (2006) | States with more unrestricted accommodations tend to have (1) higher percentages of students with disabilities participating in regular NCLB assessments and (2) lower discipline rates. |
| Gibson et al. (2005) | Some accommodations get used/recommended over others; scheduling and setting are most commonly recommended; challenges to implementation were identified; AAC Category 2 and 3 accommodations were frequently recommended and used, but caution should be taken. |
| Packer (2005) | Most common test accommodations for students with disabilities (tics) reported by parents included ET, separate location, answer recording in any way, and several others. |
| <i>Language characteristics have no disproportionate impact on Students with Disabilities</i> | |
| Shaftel et al. (2006) | Linguistic features of items have a greater effect for younger students, but no impact was found for students with disabilities. |
| <i>Educators and institutions vary in their accommodations use</i> | |
| Maccini & Gagnon (2006) | Teachers vary in their use of test accommodations (special education vs. general education); special education-trained teachers use more accommodations and number of methods course predict use. No differences in use of extended time, calculator, and read-aloud. |
| Sahlen & Lehmann (2006) | In developing policies about accommodations use, institutions need to consider their legal responsibility, the students' responsibility, the policy structure of the institution, the students' request(s) for accommodations, and the context of the course. |
| VanWeelden & Whipple (2005) | Teachers were able to administer tests with accommodations to students with disabilities (EDBD, CCD) and implement alternate assessments. |
| <i>Determining appropriate assessment accommodations is a complex and collaborative undertaking</i> | |
| Edgemon et al. (2006) | Research on accommodations can provide insight into the steps that IEP teams should follow in making decisions about accommodations. Students should be evaluated as individuals, teachers should be aware of how accommodations change the construct of interest, and accommodations should match the testing format. |
| Meyen et al. (2006) | Students with disabilities need assessments tailored to their performance level, and adaptive testing is one strategy that should be considered for the potential to lead to improved measurement for these students. |
| Rickey (2005) | The IEP team, especially special education teacher, must be recognized as responsible for making decisions regarding the education of students with disabilities. Test accommodations should exhibit a clear connection to classroom accommodations, and goals in process of identifying accommodations need to be articulated. |

Appendix G

Limitations and Future Research

Table G-1. Authors' Limitations by Study and Limitation Category

| Study | Sample characteristics | Test/Test Context | Methodology | Results |
|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| Antalek (2005) | Size and composition of sample | | | |
| Baker (2006) | (1) Homogeneity of sample may limit generalizability. (2) Missing data in data archives. | | Limitations of sample size did not allow breakdown by type of learning disability. | |
| Bolt & Ysseldyke (2006) | | Results could not be evaluated across grades due to changes in difficulty and constructs. | (1) Study design is not counterbalanced with same students. (2) There was no formal control for standardized implementation of accommodations. | |
| Bruins (2006)* | | | | |
| Cawthon (2006) | (1) Over-representation of schools for deaf and settings in South. (2) Low response rate. | | (1) Use of schools/program as unit of analysis. (2) Retrospective nature of data collection. (3) Incomplete surveys. | |
| Cohen et al. (2005)* | | | | |
| Cox et al. (2006) | Lack of reliable data from all fifty states. | | Absence of data linking performance to accommodations. | |
| Dolan et al. (2005) | | | (1) Did not address interaction hypothesis. (2) Possible novelty effect for CBT. | |
| Edgemon et al. (2006)* | | | | |
| Fletcher et al. (2006) | | | | Results are only generalizable to similar students. |
| Gibson et al. (2005)* | | | | |
| Gregg et al. (2005)* | | | | |
| Higgins et al. (2005) | (1) Small sample size. (2) Volunteer recruitment for participation: Sample potentially biased toward CBT using-schools and high SES. | Low number of passages and items. | | |

| Study | Sample characteristics | Test/Test Context | Methodology | Results |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|
| Horkay et al. (2006) | (1) Single grade. (2) Divergence from NAEP sampling frame | Only two essay tasks. | (1) Paper and CBT administration were not at the same time. (2) Differences in scorer reliability across modes. | (1) Other factors in addition to computer familiarity. |
| Horvath et al. (2005) | Small sample size. | | | |
| Huynh & Barton (2006) | No LEP or ELL students involved. | | Test was untimed for all students (thus results may not generalize to extra time accommodations situations). | |
| Kettler et al. (2005) | Only two grade levels. | Only two content areas. | Failure to operationalize accommodations and implementation. | Inability to explain why some performance was worse under accommodations. |
| Lang et al. (2005) | (1) Limited diversity of sample. (2) No knowledge of students without disabilities group variability. (3) High variability within students with disabilities group—for example, accommodations were ID'd for individuals not by disability type. | Low-stakes testing context. | | |
| Lesaux et al. (2006)* | | | | |
| Maccini & Gagnon (2006) | (1) Small sample size. (2) Unknown heterogeneity/homogeneity of classrooms. (3) Low response rate. | | (1) No way to compare respondents with nonrespondents. (2) Instructional practices list could limit responses. | |
| Mandinach et al. (2005) | (1) Small sample of disabled participants. (2) LD group could not be separated from ADHD. (3) Voluntary participation raises questions about motivation. (4) Attrition of sample. | | (1) Small numbers meant high and medium ability groups were combined, thus ability groups within students with disabilities / students without disabilities were not parallel. (2) Hard to ensure schools follow research protocol. | |
| Meyen et al. (2006)* | | | | |

| Study | Sample characteristics | Test/Test Context | Methodology | Results |
|------------------------------|--------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| Ofiesh et al. (2005) | | (1) Limited to standardized reading MC test (not essay or other formats). (2) Hard to generalize due to lack of consistency across tests used in higher education settings. | | |
| Packer (2005) | | | Limited in type of information to be collected via survey. | |
| Rickey (2005) | (1) Only exemplary schools were selected for use. (2) Small sample. | Only involved large-scale assessments, not alternate assessments. | Qualitative study, so it is descriptive and no recommendations are provided and it does not address questions about the effectiveness of particular accommodations. | |
| Sahlen & Lehman (2006)* | | | | |
| Schnirman (2005) | Low academic language proficiency in sample. | | | Floor effect. |
| Shaftel et al. (2006) | Only these three grade levels. | (1) Limited to one state's test. (2) Results cannot be generalized to other content areas. | Test item analyses could not be combined across grade levels. | |
| Sireci (2005)* | | | | |
| Sireci et al. (2005) | (1) Small, ethnically homogeneous samples. (2) Much research only involves elementary school. | | | |
| Stretch & Osborne (2005)* | | | | |
| VanWeelden & Whipple (2005)* | | | | |

* Those studies marked with an asterisk did not identify limitations.

Table G-2. Authors' Future Research Directions by Study and Future Research Category

| Study | Sample/Setting | Test/Test Context | Methodology | Results |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Antalek (2005) | | | | Study relationship between specific LD attributes and the ability to craft written language. |
| Baker (2006) | (1) Study other types of postsecondary institutions. (2) Study student groups (by age, gender, ability levels, disability classifications). | | Examine other factors that influence GPA (such as other accommodations, personality characteristics, study habits, drug/alcohol use, and social factors). | |
| Bolt & Ysseldyke (2006) | | Explore patterns to DIF and seek explanations. | | (1) Determine if read-aloud results in better measurement than if no accommodations at all are provided. (2) Understand potential relationship of other variables in impacting effectiveness of testing accommodations, including appropriateness for all students. |
| Bruins (2006) | | | (1) Track performance change in cohort over time. (2) Study effects of specific accommodations. (3) Compare state accountability workbooks. | |
| Cawthon (2006) | Diversify samples of schools and programs. | | (1) Interaction of student-, school-, and state-level characteristics. (2) Obtain more specific data on accommodations use from respondents. | Explore effect of read-aloud, signed q-r and out-of-level testing on validity and score reporting. |

| Study | Sample/Setting | Test/Test Context | Methodology | Results |
|------------------------|-------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Cohen et al. (2005) | | | | Multidimensionality suggests review of how to universally design tests. |
| Cox et al. (2006) | | | | More research into “controversial” accommodations. |
| Dolan et al. (2005) | | Involve other subject areas. | (1) Further understand effects of training. (2) Additional accommodations. | |
| Edgemon et al. (2006)* | | | | |
| Fletcher et al. (2006) | (1) Involve participants from wider age range. (2) Increase variability of reading difficulties exhibited in sample. | | (1) Assess students more thoroughly on other reading skills. (2) Focus on types of reading skills required by different tests. (3) Unpackage accommodations and evaluate in isolation. | |
| Gibson et al. (2005) | | | | Explore how IEP teams can be used to support selection/ implementation of accommodations. |
| Gregg et al. (2005) | | | Explore differences between performance of SwD and Sw/oD on specific item types. | More validity studies are needed to determine effectiveness of WJ III Cognitive Fluency cluster. |
| Higgins et al. (2005) | (1) Larger, more diverse sample. (2) Other grade levels. | Add passages and items (to improve reliability). | | |
| Horkay et al. (2006) | | | Possible unfamiliarity with NAEP laptops and variability of school computers. | |
| Horvath et al. (2005)* | | | | |

| Study | Sample/Setting | Test/Test Context | Methodology | Results |
|-------------------------|------------------------------------------------------------------------------------|---------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Huynh & Barton (2006) | | | | Examine effects of accommodations for students without disabilities. |
| Kettler et al. (2005) | | | (1) Operationalize accommodations. (2) Use single-case methods to study accommodations. | Study interaction between individual participants, tasks, and accommodations. |
| Lang et al. (2005) | Look at parental perceptions of accommodations. | Explore issues in context of high-stakes testing. | Examine perceptions of specific types of accommodations. | |
| Lesaux et al. (2006) | | | | (1) Further examine relation between word reading ability, comprehension speed, and performance. (2) Seek further insights into reading, vocab, and short-term memory under timed/ untimed conditions. |
| Maccini & Gagnon (2006) | (1) Larger samples. (2) Identify types of methods classes taken by respondents. | | Expand possible predictors list to assess reported instructional practices and accommodations. | Explore how test accommodations appropriate in type and number for students and aligned to state policies. |
| Mandinach et al. (2005) | Break out LD and ADHD participants. | Examine other tests. | (1) Randomize order of sections. (2) Include a double-time condition with section breaks for balance. (3) Obtain better/more reliable estimates of time use across sections. | (1) Examine effects of section break accommodation in isolation. (2) Research section break accommodation for functioning as intended. |

| Study | Sample/Setting | Test/Test Context | Methodology | Results |
|--------------------------|---------------------------------------------------------------------------------------------|--------------------------|--------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Meyen et al. (2006) | | | The effectiveness of CAT in assessing the performance of students with high-incidence disabilities should be researched. | |
| Ofiesh et al. (2005) | | | | (1) Clarify how test scores help justify and support need for extended time. (2) Study relationship between speeded cognitive tasks and academic tasks. |
| Packer (2005) | | | Carry out controlled research with objective measures to assess effectiveness of specific accommodations. | Examine the effects of other accommodations, besides extra time. |
| Rickey (2005)* | Focus on the variable impact of accommodations for individual students with specific needs. | | | (1) Focus on the validity of accommodations and the results obtained via their use. (2) Explore the extent to which accommodations reduce stress/anxiety for students with disabilities. |
| Sahlen & Lehmann (2006)* | | | | |
| Schnirman (2005)* | | | | |

| Study | Sample/Setting | Test/Test Context | Methodology | Results |
|------------------------------|----------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Shaftel et al. (2006) | | | Create pairs of items in word problem and computation format. | (1) Evaluate cognitive consistency of original and simplified math items. (2) Focus on the relationship between achievement in content areas and language proficiency. |
| Sireci (2005) | | Build tests not needing accommodations. | Include multiple sources of validity. | (1) Evaluate consequences of flagging/not flagging. (2) Possibly equate scores from accommodated/nonaccommodated administrations. |
| Sireci et al. (2005) | Future studies should: (1) Increase sample size. (2) Diversify samples. (3) Add in more grades. | | | (1) Research validity of interpretations from standard/nonstandard administrations. (2) Collect a variety of forms of evidence. (3) Evaluate benefits of universal test design, including technology. |
| Stretch & Osborne (2005) | | (1) Identify ways to develop tests that measure construct of interest not speededness. (2) Identify ways in test development to potentially reduce need for accommodations. | (1) Examine interaction of giftedness and timed tasks. | (1) Well-controlled valid research needed to demonstrate differential boost. |
| VanWeelden & Whipple (2005)* | | | | |

* Those studies marked with a "*" did not identify directions for future research.