

Guidebook for Studies of the Validity of Test Results for Test Accommodations

*Connecticut Enhanced Assessment Grant
Establishing the Validity of Test Accommodations
and Score Interpretations for Students with Disabilities:
A Collaboration of State-based Research*

Report prepared by
John F. Olson, Principal Investigator and Project Coordinator,
and the CTEAG Project Management Team

Guidebook for Studies of the Validity of Test Results for Test Accommodations

Connecticut Enhanced Assessment Grant

Establishing the Validity of Test Accommodations and Score Interpretations for Students with Disabilities: A Collaboration of State-based Research

*A Project of the Connecticut State Department of Education
and the
Council of Chief State School Officers
State Collaboratives on Assessment & Student Standards
Assessing Special Education Students (ASES)
Technical Issues in Large-Scale Assessment (TILSA)*

**Report prepared by
John F. Olson, Principal Investigator and Project Coordinator,
and the
CTEAG Project Management Team**

Consortium of states in CTEAG: Connecticut, Kentucky, Michigan, and Nevada

The project was funded by a grant from the U.S. Department of Education
(S368A070009)

January 2010

**Council of Chief State School Officers
Washington, DC**

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. The Council seeks member consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public.

Guidebook for Studies of the Validity of Test Results for Test Accommodations

Establishing the Validity of Test Accommodations and Score Interpretations for Students with Disabilities: A Collaboration of State-based Research

John F. Olson, Primary Author

This report is based on research funded by an Enhanced Assessment Grant from the U.S. Department of Education, awarded to the Connecticut Department of Education, in partnership with the Council of Chief State School Officers and its ASES and TILSA SCASS consortia. Publication of this document shall not be construed as endorsement of the views expressed in it by the U.S. Department of Education, the Connecticut Department of Education, or the Council of Chief State School Officers.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

Susan Gendron (Maine), President
T. Kenneth James (Arkansas), Past-President
Steven L. Paine (West Virginia), President-Elect

Gene Wilhoit, Executive Director

Council of Chief State School Officers
One Massachusetts Avenue, NW, Suite 700
Washington, DC 20001-1431
Phone (202) 336-7000
Fax (202) 408-8072
www.ccsso.org

ISBN: 1-884037-26-7

Copyright © 2010 by the Council of Chief State School Officers, Washington, DC
All rights reserved.

Acknowledgments

This research project benefited from the support and advice of members of two of CCSSO's State Collaboratives on Assessment and Student Standards: the Technical Issues in Large-Scale Assessment (TILSA) consortium, led by Doug Rindone; and the Assessing Special Education Students (ASES) consortium, led by Sandra Warren. The author wants to especially thank the project management team, the Connecticut State Department of Education staff, and the state department staff members of the validity studies for their work and participation in the project.

CTEAG Project Management Team:

Barbara Beaudin, Connecticut DOE and Project Director
Mohamed Dirir, Connecticut DOE and Project Analyst
Joe Amenta, Connecticut DOE
Gil Andrada, Connecticut DOE
Janet Stuck, Connecticut DOE
John Olson, Principal Investigator and Project Coordinator
Doug Rindone, TILSA Advisor
Sandra Warren, ASES Advisor
Martha Thurlow, Advisor/Partner
Eileen Ahearn, Advisor/Partner
Phoebe Winter, Project Evaluator
Adam Petermann, CCSSO Liaison

State Validity Study Project Leaders:

Joe Amenta, Connecticut
Gil Andrada, Connecticut
Mohamed Dirir, Connecticut
Ken Draut, Kentucky
Roger Ervin, Kentucky
Joseph Martineau, Michigan
Steve Viger, Michigan
Adam Wyse, Michigan
Carol Crothers, Nevada
Cindy Sharp, Nevada
Andrew Parr, Nevada

In addition, the author wants to thank the following state representatives for their thorough and thoughtful reviews of the draft report: Sarah Celestin, Delaware DOE; Aran Felix, Alaska DOE; Diane Henderson-Montero, Educational Testing Service; Deb Malone, Tennessee DOE; and Jana Owen, Texas Education Agency.

Acknowledgments (continued)

States participating in the Assessing Special Education Students (ASES) and Technical Issues in Large-Scale Assessment (TILSA) collaboratives, 2007–2010:

ASES	TILSA
Alaska	Arizona
Arizona	Arkansas
Arkansas	California
California	Connecticut
Colorado	Delaware
Connecticut	Dept. of Defense Education Activity (DODEA)
Delaware	District of Columbia
District of Columbia	Florida
Florida	Georgia
Georgia	Illinois
Hawaii	Kansas
Illinois	Kentucky
Iowa	Louisiana
Kansas	Michigan
Kentucky	Minnesota
Louisiana	Mississippi
Maryland	Missouri
Michigan	North Carolina
Minnesota	Ohio
Mississippi	Pennsylvania
Nebraska	Rhode Island
New Mexico	South Carolina
North Carolina	Tennessee
Ohio	Texas
Oregon	Utah
Pennsylvania	Washington
Rhode Island	West Virginia
South Carolina	Wisconsin
South Dakota	Wyoming
Texas	
Utah	
Washington	
West Virginia	
Wisconsin	
Wyoming	

Contents

Acknowledgments	1
Purpose of the Guidebook	3
Executive Summary	4
Purpose of the CTEAG Project	5
Audience for the Guidebook	5
Overview of the CTEAG Project	6
Review of Research Literature on Test Accommodations	7
Research Design and Guidelines for Conducting Validity Studies	9
Summaries of Findings from State Validity Studies	13
Overall Summary and Discussion	18
Information Provided in the CTEAG Technical Report and Project Database	25
Appendices	26

Purpose of the Guidebook

The CTEAG Guidebook serves as a practical guide for state assessment staff and others on conducting validity studies for test accommodations and interpreting the results. It provides a synthesis of the findings from the five state studies conducted for this project, with an emphasis on the use of the data and possible future applications. A companion document for the project, the CTEAG Technical Report, provides more details of the studies, methodology, data analyses, and other technical information.

Executive Summary

The No Child Left Behind Act requires that states offer accommodations on grade-level assessments in order to provide tests that are accessible to as many students with disabilities as possible. However, little research has been conducted on the validity of accommodated score interpretations or the effectiveness of test accommodations. From 2007–2009, with Connecticut serving as the lead state and with the support of the Council of Chief State School Officers (CCSSO), a consortium of 38 states who are members of two of the larger State Collaboratives on Assessment and Student Standards (SCASS) groups—Assessing Special Education Students (ASES) and Technical Issues in Large-Scale Assessment (TILSA)—participated in a special research project called the Connecticut Enhanced Assessment Grant (CTEAG). The project was funded by a grant from the U.S. Department of Education (S368A070009). The research conducted by CTEAG provides empirical evidence to help establish the validity of inferences from accommodated tests based on specific student accessibility needs.

The project was coordinated across four volunteer states (Connecticut, Kentucky, Michigan, and Nevada). A rigorous research design was used, with three states using the same 2 x 2 counterbalanced research design in which students (in general education or special education) were crossed with accommodations (with or without) on parallel split-half test forms administered in their state. One state, Michigan, used an alternate design to study its alternate assessment based on modified achievement standards. Guidelines for planning the accommodations validity studies were provided to the participating states to assist them in preparing and organizing key activities for conducting the studies. Detailed information was developed and shared on five main steps for the studies: design, test forms, samples, test administration, and data files. Each volunteer state selected an accommodation to study and conducted at least one validity study (Connecticut conducted two), with most states selecting read-aloud accommodations. Between 300 and 700 students were tested in each state's sample. Analyses done by project staff included statistical analysis of items and test scores, exploratory and confirmatory factor analysis to examine for any structural changes in test constructs due to accommodations, and repeated measures analysis of variance to examine main and interaction effects.

In four of the studies, it was found that, for the most part, test accommodations did not inappropriately alter the item statistics or the overall test characteristics. Additionally, accommodations changed neither the dimensionality of the tests nor the test structure. In three of the studies, accommodations increased scores for students with disabilities significantly more than students without disabilities.

Deliverables from the project include this guidebook, a separate CTEAG technical report, and an associated database that provides student data, procedures, statistical data, and other information for evaluating the validity of test results from accommodated assessments. The findings and products can be used by states, where applicable, in providing evidence on the appropriateness of certain accommodations that are used in their state assessments.

Purpose of the CTEAG Project

The main goal of the CTEAG project was to conduct in-depth studies of the validity of score interpretations from selected test accommodations used for students with disabilities. The project's purpose was to empirically examine the effects of accommodations on test performance, comparing students with disabilities (SWD) to students without disabilities (SWOD) and accommodated with non-accommodated test administrations, and to look at the impact, if any, of accommodations on the psychometric characteristics of the items and tests. The products of the project include a guidebook summarizing the findings from the multiple validity studies that were conducted, along with a technical report and an extensive database of information from the research, so that all states can use and refer to the evidence when justifying the use of certain accommodations and the validity of their accommodated assessments and interpretation of scores.

Because the final reports and database include information on research designs, procedures, and findings from the five validity studies, they are a valuable resource for every state that needs to reference existing data and evidence to inform, evaluate, and support their decisions on the use of various types of accommodations. CTEAG project documents and materials can be found on the CCSSO website (<http://www.ccsso.org>). Also, because the results from the states cover a variety of different read-aloud accommodations (for example, read alouds done by the teacher or trained test administrator and read alouds delivered by computer), the project provides a resource that few, if any, states would be able to develop by themselves. This compilation and sharing of information can be a great benefit for states in meeting NCLB peer review requirements for standards and assessments by providing additional validity information concerning inferences about student achievement and the performance of students with disabilities.

In addition, information from the guidebook can be used by state staff and others who have decided to conduct validity studies on test accommodations, and for those who are thinking about possibly conducting such a study in the future.

Audience for the Guidebook

The intent of the guidebook is to be accessible to a wider audience than only assessment personnel, psychometricians, and researchers. The intended audience includes all people in state education agencies, local districts, and special education, as well as policymakers and other key stakeholders. Thus, as much as possible, given the nature and content of the report, it has been written in a non-technical manner.

Overview of the CTEAG Project

The lead state for the CTEAG project was Connecticut. With the support of CCSSO a large consortium of states from the ASES and TILSA SCASS groups participated in this special research project, conducted between October 2007 and February 2010 (a list of all participating states in the two SCASS groups is shown in the Acknowledgments). The combined strength of this consortium brought together the expertise and practical experience of state personnel who are working with state assessments, special education, accommodations decisions, technical issues, and validity evidence to support their programs.

During the course of the CTEAG project, the Connecticut State Department of Education (CSDE) coordinated closely with CCSSO to manage the project's work and the subcontracts with specific researchers. Work on the project was directed by a Project Management Team (PMT) and included experts in assessment, special education, and technical issues. State members from the TILSA and ASES SCASS groups led the validity studies in their states. A list of the organizational structure and key personnel in the project is provided as an appendix.

A detailed master schedule was developed for the duration of the CTEAG project. The timeline ran from October 2007 through February 2010. The work of the CTEAG was organized around the regularly scheduled meetings of the ASES and TILSA SCASS groups, when CTEAG project members would come together for meetings and interactions. A detailed version of the master schedule can be found in the technical report.

Summary of Approach Used for CTEAG Project

As described earlier, the purpose of this project was to conduct research on the validity of various types of commonly used test accommodations. Validity was evaluated by comparing their effectiveness for SWD, who require the accommodation, to their effectiveness for SWOD, who do not typically use the accommodation. The specific impact of the use of accommodations on test scores was analyzed. (More details on the research design, methodologies, and analysis procedures are included in a subsequent section of this guidebook. Additional details also can be found in the CTEAG Technical Report.)

- The lead state for this grant was Connecticut.
- The other states that volunteered to conduct validity studies were Kentucky, Michigan, and Nevada.
- A project management team comprised of CSDE staff, researchers, experts in the field, and representatives from CCSSO and states oversaw and coordinated all project activities.
- CCSSO provided logistical support, key staff, and consultants to assist with activities for the project.

- The TILSA and ASES SCASS groups collaborated on the project. TILSA provided the technical and measurement “thought leadership” and ASES provided the special education content and process “thought leadership” for this effort.
- The individual states that volunteered to conduct validity studies each carried out a study in their state, assisted in the development and administration of the specific accommodated assessment, and provided data to the project’s statistical analyst.
- CTEAG project staff conducted all analyses, summarized the results, and wrote the reports for the project.

Review of Research Literature on Test Accommodations

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) provides important recommendations for documenting the validity of assessment accommodations. These are very similar to NCLB requirements for technically sound assessments, as described in the legislation and peer review guidance. Chapter 10 of the joint standards provides background information on testing individuals with disabilities and addresses many issues related to accommodations, strategies of test modification, and issues in reporting scores. An underlying principle in the standards is ensuring the validity of the assessments. As stated in Standard 10.1: “In testing individuals with disabilities, test developers, test administrators, and test users should take steps to ensure that the test score inferences accurately reflect the intended construct rather than any disabilities and their associated characteristics extraneous to the intent of the measurement.”

The standards 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). Over 70 different accommodations in eight categories (i.e., motivation, assistance prior to testing, scheduling, setting, directions, assistance while testing, use of equipment/adaptive technology, and changes in format) have been identified in the testing literature (Elliott, Kratochwill, & Schulte, 1998). Among the issues at the forefront of conversations in many testing contexts today are how states will report progress for students with Individualized Education Programs (IEPs) as part of NCLB, what accommodations are used, how they are implemented, and to what extent scores are comparable from tests administered under standard and non-standard administrations.

Assessment Accommodations: An Overview of Recent Research Approaches

The body of research on test accommodations has continued to grow over the years, although it is still relatively small. A summary of recent research on the topic can be found in an online bibliography from the National Center on Educational Outcomes (NCEO) (<http://www.cehd.umn.edu/NCEO/Things%20to%20delete/oldAccomStudies.htm>). As Martha

Thurlow stated in a recent American Educational Research Association paper summarizing research findings on accommodations (2007, p. 2): “There are many definitions of testing accommodations, yet all now converge on the same concepts. In general, any type of change to testing materials, setting, or procedures that does not alter what is being measured is considered to be an accommodation.” Some researchers refer to the elimination of construct irrelevant variance, in other words, the variance associated with extraneous features of test administration (Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000). Demonstrations of test changes that maintain the validity of the construct that an assessment is attempting to measure have been challenging for the field, and numerous approaches to research have been pursued (Thurlow, McGrew, Tindal, Thompson, Ysseldyke, & Elliott, 2000; Sireci, Scarpati, & Li, 2005; Tindal, 1998), including single subject designs, “boost” studies, and “differential boost” studies. (More details on differential boost studies are provided later in the guidebook.)

Accommodations continue to be a topic of heightened interest for a number of reasons. Participation in national, state, and district assessments is required by both the federal special education law, the Individuals with Disabilities Education Act (IDEA), and by the federal Elementary and Secondary Education Act (ESEA), reauthorized and amended as the No Child Left Behind Act (NCLB), which requires participation in state assessments and the National Assessment of Educational Progress (NAEP). First and foremost, it is no longer a question of whether students with disabilities need to be participants in national, state, and district assessments. Many students with disabilities require accommodations; data from the 2003–2004 Annual Performance Reports submitted by states to the federal government indicated that an average of 61–66 percent of school-age students with disabilities used accommodations during assessments (Thurlow, Moen, & Altman, 2006).

A major challenge surrounding accommodations is determining which changes to test materials, settings, and procedures result in alterations to the intended construct. Historically, this concern has not always been a top priority. For some time, the major focus was on the participation of students in the assessments—accommodations were seen as an avenue to participation, and the distinction between whether a change compromised the construct being assessed was not necessarily the major factor of interest (Thurlow, 2007). It has taken time for the field to adjust its focus to clearly distinguish between accommodations and modifications (those changes in materials, setting, or procedures that result in changes to the construct measured), and for research and policy to address these concepts.

NCLB has helped to increase the focus on the use and validity of accommodations. As Zenisky and Sireci stated in their report (2007), NCLB has had the effect of “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.” Their report provides an update on the state of the research on testing accommodations and identifies promising areas of research to further clarify and enhance understanding of current and emerging issues. While many of the studies reported that the use of accommodations had some positive effect on test scores, it was

noted that variations across studies in the operational definitions of those accommodations does challenge the extent to which findings can be generalized across studies. Furthermore, while much work has been 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.

In the recent book *Large Scale Assessment and Accommodations: What Works?* Cara Cahalan-Laitusis of Educational Testing Service authored a chapter on the variety of approaches for studying accommodations on assessments (Chapter 6, *ibid*, 2007). In her chapter, she provides policymakers, researchers, and test developers with a summary of research-based approaches for determining the validity and reliability of test scores obtained under accommodated conditions. Cahalan-Laitusis focuses on determining if the outcome of an accommodation used on a particular assessment yields valid test results and if it removes construct irrelevant variance to improve the validity of the test score. She argues that the differential boost model, or the interaction model, may be a preferred approach to use in interpreting the effects of accommodations. Differential boost indicates that SWD receive significantly larger gains in overall test score from an accommodation than SWOD and, although both groups can receive improved test scores by using the accommodation, the gains for SWD are differentially larger. The interaction model has a similar premise, but methodologically uses a repeated measures analysis of variance to examine for a significant interaction between disability classification and accommodation when all students have taken the test under both accommodated and non-accommodated conditions.

This is the approach that CTEAG used in its studies—an experimental research design that examined differential performance boost using an interaction model. Also, the intent of the CTEAG project was to build upon the existing body of research, expanding the body of knowledge that currently exists on test accommodations and transferring principles found to be successful in strengthening the validity of state assessments. Indeed, given the policy emphasis that has been placed on the performance of subgroups of students in state and national assessments, 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.

Research Design and Guidelines for Conducting Validity Studies

In this section, a description of the research design, in general terms, is provided so that users of the guidebook will be able to interpret results from the validity studies. Information in this section includes a brief mention of the requirements for the studies and the guidelines provided by the project to support states in conducting their studies.

Research Design and Procedures: Each of the states was guided to use the same 2 x 2 counterbalanced research design that had students (in special education or general education)

crossed with accommodations (with or without) on the test forms used in their state (see the following table). The intent of the project was to have a variety of different accommodations, grades, and content areas selected for the studies so that validity evidence was gathered under a variety of conditions for the project as a whole. Each of the participating states first decided what accommodation to study and then determined the content area and grade level to administer.

Accommodations		Students	
		Special Education	General Education
Without	Split half A1	Split half B1	
With	Split half B1	Split half A1	

Instruments: Each state was responsible for creating a pair of abbreviated test forms for use in the study. In order to reduce the overall testing time on students, two parallel split-half forms were developed from a full form, and each subform was administered with and without accommodations, with each student receiving one subform with accommodations and the other without. The forms were based on an operational test, or possibly, a released form. The choice of using current or old items/forms was left to the state. To reduce the burden on states, no new item development was required for the project. Each subform consisted of half the items from the full test. These short versions of an existing test were made by splitting the full form into halves that were parallel in content and difficulty, based on pre-existing item parameters if they were available. Working with their vendor, state content staff in the mathematics or reading subject areas, along with the state’s psychometrician, helped create these forms and/or conduct reviews to ensure that the forms were indeed parallel, both statistically and in content. When possible, pre-calibrated item parameters from previous administrations of the items were used in form construction.

Sample: Based on the accommodation selected for its study, each participating state identified the students who required that specific accommodation. The characteristics of the sample of students with disabilities depended upon the accommodations used. For example, in a state studying a read-aloud accommodation, all students in the tested grade level whose IEP specified a read-aloud accommodation were eligible for the study. The majority of these students might have a learning disability in reading, but other disabilities such as vision impairment could be represented in the eligible pool. Depending on the size of the eligible population, states included either a sample or the entire pool of eligible students with disabilities in the study. Generally, a minimum of 150–200 SWD were needed in each state’s sample. Once the special education sample was identified, a similarly sized sample of general education students was selected to

participate in the study, matched on their general demographics. General education students were randomly selected from the same schools or classrooms if appropriate.

Administration: After the subforms were created, they were printed by each state's vendor to be similar in appearance to the regular state assessment and then shipped to the schools. Each state determined the best time frame for the administration of their validity study. These were typically done as a separate, special administration and not in conjunction with the operational administration of the state program, since it was important that this research effort not adversely affect the operational administration of the state assessments used for accountability and reporting of Adequate Yearly Progress (AYP). Each student was administered both subforms used in the study, one subform at a time, receiving one under standard conditions and the other with accommodations (or vice versa). The administration of the subforms was to be counterbalanced to control for any possible order effects. In other words, the presentation of forms was supposed to be equally balanced by accommodation condition so that half the students were tested with an accommodation first and the other half tested with the accommodation last. In addition, each subform was administered in both standard and accommodated conditions so that there were an equal number of responses in each condition per form. After the administration, the state's contractor scanned and scored the student responses and then sent a data file to the CTEAG project analyst, who conducted the statistical analyses for each state. Prior to the transmittal of this data, the project analyst worked with each state to agree on a format layout of the data files to be sent.

The information on the following page was provided to the participating states to assist them in preparing and organizing key activities for conducting the validity studies. Detailed guidelines were developed on five main steps for the studies: design, forms, sample, administration, and data. These guidelines were prepared by the project analyst and principle investigator, reviewed by the CTEAG PMT, and then shared with the states and discussed individually with each one.

Validity Analyses and Statistical Methodology

The purpose of the studies was to examine the validity of test scores for the accommodated administration, and the primary focus was related to examining content and construct validity and the psychometric characteristics of the tests. Validity investigations addressed three important areas: 1) item and test characteristics, 2) test structure, and 3) effects on performance for students with disabilities (SWD) in comparison to students without disabilities (SWOD).

**Guidelines in Planning for the Accommodations Validity Studies:
Steps for States to Follow in Preparing to Conduct a Study***

Select the subject area, grade, and type of accommodations to study

Each state will be testing only one grade and one subject (mathematics or reading). Select the type of accommodation you want to study. Make sure you have a large enough number of students for the type of accommodation you want to study.

Build the project test forms for the study

Use an existing form of the state test (such as a released form, alternative form, breach form, etc.). Divide the form into two parallel half-forms. Have content staff and psychometricians review and approve the parallel forms. Arrange production, printing, shipping/receiving, and scoring of the test with the test vendor.

Sample students

A minimum of 150–200 students is needed for each of the two groups: a students with disabilities (SWD) group and a students without disabilities (SWOD) group. After you are satisfied with your selection of the SWD group, select a sample from the SWOD population. As much as possible, this sample should match the demographics of the SWD group.

Test administration

The state will need to provide written test administration instructions that are unique to this study for test examiners. Be as flexible as possible in determining the best schedule for conducting the study in light of the state's testing cycle. Allow students to become familiar with the accommodation by using it at least once prior to the test administration and then again when directions for administration are given on the day of the assessment.

Data

Project staff will conduct all analyses. The main data to be analyzed for the study are the student scores from the test administration. After the student responses have been scored, provide a data file from your test vendor with all student-level data, including raw data, scoring keys, item scores, student ID, and other information.

* The complete set of guidelines is provided in the Appendix. The ones shown here are an abbreviated set that summarizes each of the five steps.

The research questions examined the following:

1. Does the accommodation affect items statistics?
2. Does the accommodation change the construct and structure of the test, i.e., is there an impact on the factor structure under an accommodated administration?
3. Does the accommodation increase performance of students? Does it increase performance more for SWD than it does for SWOD?

To answer these three questions, a variety of investigations and statistical analyses were performed, as described below:

- ❑ Investigation # 1: Impact of the accommodation on the statistical properties of the items and the total test form.
 - Compared classical item statistics (difficulty, discrimination) between administration conditions.
 - Compared test form reliabilities.
- ❑ Investigation # 2: Impact of accommodation on the underlying test construct and structure.
 - Examined factor structure and dimensionality using exploratory and confirmatory factor analysis.
- ❑ Investigation # 3: Impact of accommodation on test scores for the two groups of students.
 - Conducted a repeated measures analysis of variance (ANOVA) with standard test as pretest to investigate differences in main and interaction effects.

In regard to this last question, a repeated measures ANOVA was used to examine the effects of the accommodations on student performance, as measured by test scores. This statistical procedure examines whether accommodations significantly increased test scores for one or both of the groups, and if it was more for SWD than for SWOD. As such, a 2 x 2 counterbalanced design was used for the two groups of students (SWD and SWOD) and two administration conditions (with and without accommodations). Since each of the four subsamples took two subforms that were designed to be parallel, main and interaction effects were examined by comparing differences among the scores, the significance of the differences, and the effect sizes of the differences by conducting a 2 x 2 repeated measures ANOVA.

Summaries of Findings from State Validity Studies

In this section, an overall summary of the five validity studies is presented with information on each state’s procedures and the findings from its study. More details from each study are provided in the technical report for the CTEAG project.

The following table provides an overview of the studies and includes information such as the accommodations used, content areas and grades involved, and when the test administration occurred. As shown in this table, in four of the five studies, states decided to study a read-aloud accommodation. However, states used several different approaches for this. Connecticut used a computerized text reader to deliver the accommodation, with a digital voice for a reading comprehension test for their first study and with a recorded human voice for a mathematics test for their second study. Kentucky and Nevada used trained human readers for the read-aloud accommodations, but in different grades and subject areas—Kentucky in reading at grade 4 and Nevada in mathematics at grade 7. Michigan compared the results from a pilot version for its alternate assessment based on modified achievement standards (AA-MAS) to the general test. The state viewed the enhanced directions/scaffolding as an accommodation for its AA-MAS. Four states administered their studies at a variety of grade levels in the areas of mathematics and reading, while Michigan targeted reading and writing in their English language arts assessment.

Summary of General Characteristics of the Five State Validity Studies

Topics/Issues	Connecticut # 1	Connecticut #2	Kentucky	Michigan	Nevada
Accommodation	Computerized text reader with a digital voice	Computer-delivered with a recorded human voice	Read-aloud by a trained human reader	Enhanced directions/scaffolding and fewer response options	Read-aloud by a trained individual (e.g., teacher or aide)
Grade	7	5	4	6	7
Content Area	Reading Comprehension	Math	Reading	English Language Arts	Math
Administration: <ul style="list-style-type: none"> • Test Date • Study Design • Items/Forms Development 	May 2008. Computer-based test forms with 20 MC items per form.	Spring 2009. Computer-based test forms. Used old secure math test with 26 MC items per form.	Mar. 2009. Developed forms from a pool of non-state items that were similar in content to KY forms, with 22 MC items per form.	Oct. 2008 and Jan. 2009. Used released items (24 MC and 1 CR). Data from MEAP-Access for SWD and MEAP admin done earlier.	Nov.-Dec. 2008. Used 20 MC items per form.
Vendor	Measurement Inc.	Measurement Inc.	Measured Progress	Measurement Inc.	Measured Progress and WestEd
State contact(s)	B. Beaudin M. Dirir G. Andrada J. Amenta J. Struck	B. Beaudin M. Dirir J. Amenta G. Andrada J. Stuck	K. Draut R. Ervin	S. Viger J. Martineau J. Griffiths A. Wyse	C. Crothers C. Sharp A. Parr

Because of timing issues and state schedules for their regular ongoing state assessment programs, the administration dates for the studies varied somewhat. Validity studies were conducted in a time frame that ranged from May 2008 to May 2009. Two states administered the studies in the fall-winter timeframe (late 2008 to early 2009). All states conducted their administrations at times that did not conflict with their regular state assessments.

In Chart 1, general information is provided on whether states met two important criteria for conducting the studies: use of the 2 x 2 research design and adequate sample size. All states except Michigan followed the design that was specified in the guidelines. Michigan used an

alternate design that was approved for the study after a series of discussions with the CTEAG PMT. Their results were included because, despite the change in the design, their study still provides important information about the use of accommodations in NCLB-like tests. The design utilized a pilot version of an assessment based on the general state assessment, the Michigan Education Assessment Program (MEAP), in preparation for a new state test, the MEAP-Access. At the time of the study, the MEAP-Access was being designed as an AA-MAS developed by the state under NCLB regulations. For this validity study, a pilot form was developed to be parallel to the MEAP in content using released items from the test that had one answer choice removed and was accompanied by enhanced directions. Students in the study took the items from the two forms at separate times and the forms did not share any items (i.e., there were no common items between MEAP and the pilot form). Because of this, some of the analyses could not be performed (as described further below).

All states met the minimum sample size requirements of at least 150–200 students in each subgroup.

Chart 1. General Information on Requirements Criteria for Studies

General Information	Connecticut # 1	Connecticut # 2	Kentucky	Michigan	Nevada
2 x 2 Design Followed?				Alternate Design Used	
Adequate Sample?					

In Chart 2, the findings from the analyses for the three research questions are presented. For the first research question, the item difficulties and point-biserial correlations (a measure of discrimination) were examined. In general, the item statistics were unchanged across the two testing conditions, except in the case of Kentucky, where items had higher p-values (percent correct) under the accommodated condition (i.e., students performed better, and values for some point-biserial correlations decreased). Because no items were common across testing conditions in Michigan, the effect of the accommodation on individual item statistics could not be analyzed.

For the second research question, the exploratory and confirmatory factor analyses of test structure found no changes in dimensionality or factor structure across the five studies.

Chart 2. Summary of Findings from CTEAG Validity Studies

Accommodation Effects	Connecticut # 1	Connecticut # 2	Kentucky	Michigan	Nevada
Item Statistics Not Changed			Changes in item stats	Not applicable	
Test Dimensionality Not Changed					
Test Structure Not Changed					
Significantly Increased Test Scores					 **
Significantly Increased Test Scores for SWD More Than SWOD (Disability X Accommodation)					

**For Nevada study, $p=0.0523$

For the third question, the analyses of main and interaction effects found that the accommodations significantly increased test scores for students in two of the five studies (Kentucky and Michigan). Analysis of the interaction effects found that scores for SWD increased more than scores for SWOD in three of the studies (Connecticut #2, Kentucky, and Michigan). In Connecticut study #2, scores for SWD increased slightly, whereas scores for SWOD decreased under the accommodated condition. In both Kentucky and Michigan, scores for both groups increased, with those for SWD increasing much more than those for SWOD. Plots of the interaction effects for these three studies are shown in the following figures.

Figure 1. Connecticut #2—Plot of Means for Disability Status by Accommodation Condition

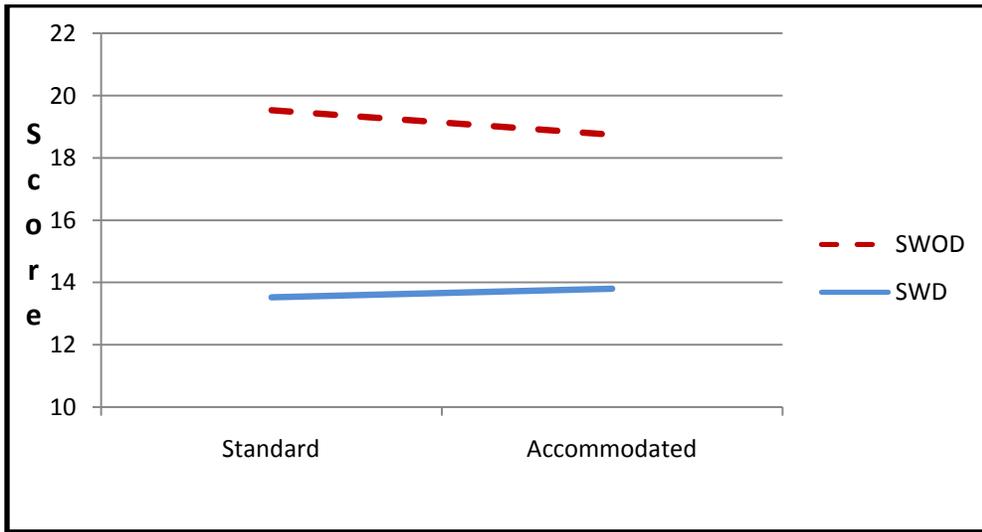


Figure 2. Kentucky—Plot of Means for Disability Status by Accommodation Condition

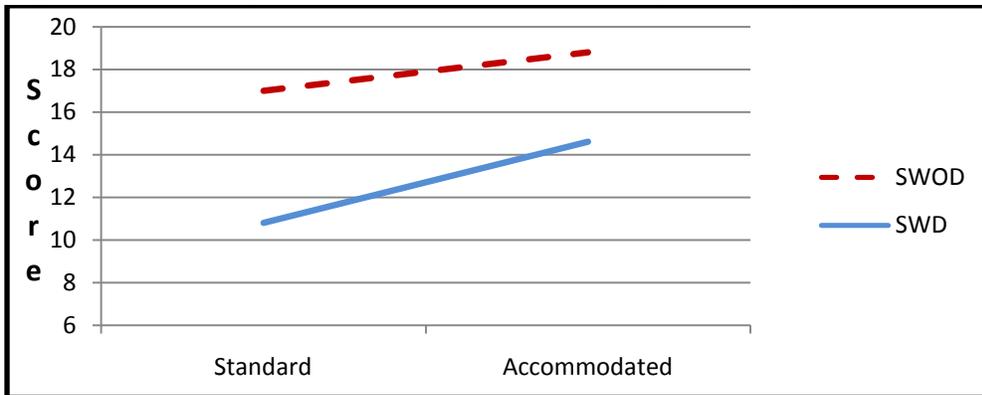
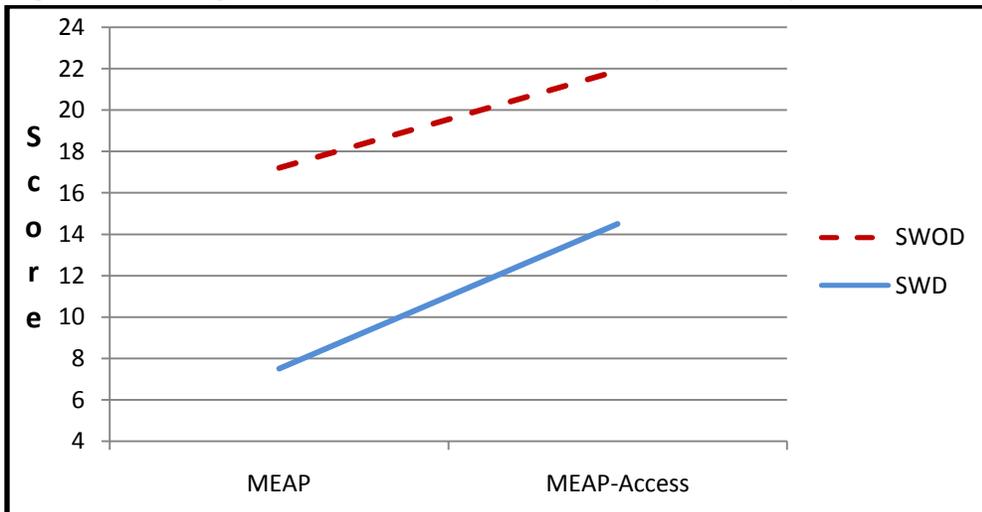


Figure 3. Michigan—Plot of Means for Disability Status by Accommodation Condition



A Discussion of Differential Boost

In earlier years, researchers studying the effects of accommodations proposed that for the accommodation to support valid inferences, it should have a positive effect on scores of students with disabilities and no effect on the scores of students without disabilities (Phillips, 1994). If the accommodation increased scores for both groups of students, the general conclusion at that time was that the accommodation provided an unfair advantage to some of the students who used it. This reasoning was based on the idea that students without disabilities did not have access needs that were addressed by the accommodation and any increases in their scores were probably due to changing the item difficulties or the construct targeted by the test.

However, in more recent years, research has progressed and the paradigm has shifted to one of “differential boost” and in particular, research on student access to test content and opportunity to perform (Cahalan-Laitusis, 2008). Researchers have proposed that an accommodation may support valid inferences if students with disabilities have a relatively greater increase in scores with the accommodation than students without disabilities. Just because the scores of students without disabilities also increase does not automatically invalidate the accommodation (Sireci, Li, & Scarpati, 2005). Recent research has acknowledged the fact that students who are not identified as having a disability may still have access needs that are alleviated by the accommodation. For example, many poor readers are not identified as having a communication-related learning disability, but they may still benefit from an accommodation in which all or part of a test is read aloud to them. It is especially important to take this perspective into account when reviewing the findings from the CTEAG validity studies.

Overall Summary and Discussion

In this section, a synthesis of the findings across the five validity studies is presented along with a discussion and interpretation of the results in light of their practical use by states and others. In order to make fair comparisons across the studies, the overall findings for Connecticut studies #1 and #2, Kentucky, and Nevada will be discussed first. The findings for Michigan will be discussed separately since that study was conducted under different conditions and focused on the use of a pilot form of an AA-MAS.

More details and additional information for each of the state studies are included in the CTEAG technical report.

Research Question 1: Does the accommodation affect item statistics?

Overall, as was shown in Chart 2, the various test accommodations had little effect on item statistics. In the two studies conducted by Connecticut and the one by Nevada, the difficulties and discrimination indices did not change, nor did the reliabilities of the test forms.

Thus the use of read-aloud test accommodations did not seem to have any impact on or change the item and test characteristics. However, for the Kentucky study, several of the items became easier and the point-biserial correlations decreased under the accommodation condition, with some items' point-biserials decreasing dramatically. As a result of the decrease in test score variability, the reliabilities of the forms used in Kentucky also decreased slightly under the accommodated condition. A possible interpretation of this effect is that the read-aloud accommodation used with the Kentucky reading assessment made the items in the test much more accessible and thus easier for students, as well as somewhat less discriminating because of less variability in the scores.

Research Question 2: Does the accommodation change the content structure of the test, i.e., is there an impact on the factor structure under accommodated and standard administrations?

In general, the use of test accommodations did not seem to have any effect on the test constructs for the various assessments that were part of the validity studies. Based on the results from the exploratory and confirmatory factor analyses, none of the read-aloud accommodations changed the dimensionality of the tests or had an impact on the content structure (i.e., test constructs). Similar factor structures were found for both the accommodated and non-accommodated test forms. Thus, based on the findings from the validity studies conducted for this research project, it is reasonable to say that when comparing accommodated to non-accommodated (standard) test administrations, the use of read-aloud accommodations had no impact on the test constructs.

Research Question 3: Does the accommodation increase performance of students? Does it increase performance more for students with disabilities than it does for students without disabilities?

In answer to this question, the results from the validity studies are mixed. The findings for the analysis of main effects from the repeated measures ANOVA found that the read-aloud accommodations significantly increased test scores for all students (SWD and SWOD combined) in one of the studies (Kentucky). The accommodations did not significantly increase scores for students overall in the two Connecticut studies or in Nevada (although scores did increase in Nevada, but not quite enough to be statistically significant). For the former two studies, this may have had something to do with the mode of test administration used in Connecticut (i.e., use of a computerized read-aloud and students overall lack of familiarity with it). The significant main effects found for Michigan are discussed below.

However, besides examining main effects, the possibly more interesting research question is whether there was a differential boost effect from the test accommodations. Again, the results were mixed. Analysis of the interaction effects found that scores for SWD increased more than scores for SWOD in two of the four studies (Connecticut #2 and Kentucky). This

finding provides support in the positive benefits of using the read-aloud accommodations to make the test more accessible for SWD and thereby improve their performance. In addition, some SWOD in these two states also benefitted from use of the accommodations, which is a reasonable outcome when one considers that some students who do not have a disability may have access needs that also are helped by the read-aloud accommodation. Note that no significant interaction effects were found for the Connecticut #1 or Nevada studies. The significant interaction effects found for Michigan are discussed below.

Discussion of Results from Michigan’s Validity Study

As described earlier, the validity study conducted by Michigan used a slightly different approach from those of the other states in the CTEAG project. While the other studies in this project focused on accommodations to the general test so that students were still assessed against the same grade-level achievement standards as used for the test without accommodations, the Michigan study looked at a pilot form for a future alternate assessment based on modified achievement standards (AA-MAS). The MEAP-Access will measure the same grade-level content expectations as does the MEAP, Michigan’s general assessment. While both accommodations to the general test and the AA-MAS are designed to increase access for students with disabilities, accommodations to the general test allow students to be assessed based on general grade-level achievement standards, whereas the AA-MAS allows students to be assessed based on achievement standards that are typically easier to meet than the general achievement standards.¹

In late 2008, as Michigan was beginning the process to design the new MEAP-Access assessment for use as an AA-MAS, the state had a unique opportunity to evaluate the validity of enhanced directions on the English Language Arts portion of their assessment within the framework of the CTEAG study. For the pilot form, MEAP items were revised to make them more appropriate for students with disabilities, which included the use of enhanced directions/scaffolding and fewer options for students to choose from on multiple-choice questions. For the purposes of the study, the enhanced directions were viewed as an accommodation. The use of enhanced directions with the pilot MEAP-Access form allowed the test administrator to read and clarify parts of the test to help students better access the reading and writing portions of the assessments.

The objectives of the Michigan validity study were to examine whether the use of enhanced directions alters item properties and test structure, and whether the revisions affect test scores for SWD differently than test scores for SWOD. However, due to lack of common items between the two forms that were studied (MEAP and pilot MEAP-Access) it was not feasible to assess the effects of the enhanced directions/scaffolding on individual test items since none of

¹ According to the U.S. Department of Education (2007), “A modified academic achievement standard is an expectation of performance that is challenging for eligible students, but may be less difficult than a grade-level academic achievement standard” (p. 14).

the items were used under the two conditions—once in the regular administration and once in the AA-MAS condition. Looking at students’ performance on the test forms in terms of average difficulty, it was found that average performance on the MEAP-Access form was higher than on the MEAP form. Thus, as expected, the changes made the set of test items on the pilot form less difficult, on average, than their original iteration. The results also indicated that the SWD group benefitted more from the enhanced directions/scaffolding condition than the SWOD group did.

For interpretation purposes, the factor structure of the two forms was found to be similar (i.e., the changes did not affect the test construct being measured). Both forms were found to have a single underlying factor, which could be interpreted as an English Language Arts factor. The examination of the main and interaction effects on test scores revealed that the use of enhanced directions had a significant and substantial effect on increasing test scores, and the interaction effect was significant, with SWD differentially gaining more from the accommodation than SWOD did. Although both SWD and SWOD showed increases in their test scores, there was a differential boost and the effect was much larger for SWD. Therefore, based on the results from these analyses, it appears that the pilot MEAP-Access with enhanced directions/scaffolding used in Michigan had a major impact on increasing the accessibility of the test and on increasing the scores of students, especially benefitting the SWD group.

Study Limitations and Areas for Further Research

Throughout the course of the CTEAG project, the researchers noted various issues where methodologies and approaches could be addressed further and possibly be improved. In this section, it is important to discuss possible problem areas that may affect interpretations of the results, with the expectation that, if possible, these areas be addressed in future research.

First, the results from this collection of validity studies are specific to the accommodations that were used and the contexts in which they were applied. The majority of studies examined the effects of read-aloud types of accommodations, so it is not necessarily possible to generalize the results that were found to all other accommodations.

Carrying out research on students with disabilities is often challenging for a number of reasons. For example, obtaining adequate sample sizes for SWD in order to conduct statistical analyses can be difficult, if not sometimes impossible. Many of the techniques typically used to analyze test data are methodologies that require large samples—for example, factor analysis or structural equation modeling usually requires sample sizes of at least 300 or more. Other statistical approaches usually require at least 100 students per group and preferably more to have enough power to detect small effect sizes. However, in some states it may be impossible to obtain sample sizes of even 100 for low-incidence disabilities within small populations of students. For example, the number of students who use a Braille accommodation at a given grade level may be less than 10 in an entire state. Although the CTEAG project required states to select accommodations for their validity study only if there were at least 150 students in each group, the results from the studies may have been stronger if the sample sizes had been larger. In

addition, other types of analyses that are useful for comparing the performance of subgroups of students could have been done with larger sample sizes, for example, conducting a Differential Item Functioning (DIF) analysis to evaluate for item-level differential performance.

Another area that needs to be considered when doing these types of validity studies is that the SWD samples are often quite heterogeneous. There is much variation in any type of disability and the classification of students into disability categories may cover a broad definition of that disability. The natural variability in the student samples needs to be taken into account when planning for a study and in interpretations of the findings.

In order to interpret the results more fully, CTEAG attempted to collect data on the specific types of disabilities that students in the state samples had. However, many states did not have this information in their data files and some do not collect it as part of their assessment programs. So, this was not possible. This is an area in which more data on SWD would be useful for researchers.

A related issue is that many students, such as students who have visual or auditory disabilities, use a bundle of accommodations during the test administration. This makes the study of a single type of accommodation problematic if the student is used to getting a package of accommodations, for example a read-aloud in an individual administration with extended time. Isolating the impact of just one accommodation and disentangling the effects on the validity of test scores can be difficult. This issue was discussed in depth by the CTEAG PMT, and unfortunately there was no simple answer. In discussions with the states, it was decided to study a single accommodation in each of the studies, given the realities mentioned above. In the future, it would be beneficial for researchers to study the effects of combinations or bundles of accommodations on test validity, as the results would be more comparable to what actually is being done in the classroom and on state assessments.

Finally, researchers have learned that an important consideration to be taken into account is determining more clearly which students benefit from an accommodation that is intended to increase access and provide students with the opportunity to demonstrate their knowledge, skills, and abilities. Follow-up analyses of large-group data and studies that include measures of student access needs will help us better provide accommodations that support valid inferences. In addition, more study of the provision of accommodations to students is needed—for example, there is little research about the degree to which the accommodations selected by IEP teams are the accommodations most appropriate for increasing access.

Practical Applications and Implications for Use of Results from CTEAG Validity Studies

The guidebook has provided a summary of the research design, procedural guidelines, and methodologies used for the collection of validity studies on test accommodations, as well as a synthesis of the findings from the five state studies and discussion of the results. As was stated at the beginning of this report, the purpose of the guidebook is to serve as a practical guide for state assessment staff and others on conducting validity studies for test accommodations and

interpreting the results. In this section, practical use of the data and possible future applications are emphasized. Because the research can provide information at a variety of levels, this section also focuses on the implications for the use of the results for the following groups: states, teachers, policymakers, and researchers.

States: One of the most important uses of the information from the CTEAG project is to help states refine their decision making procedures with respect to the appropriate accommodation of students with disabilities on the general assessment and the appropriate inclusion of students in the alternate assessment. Accommodation decisions need to be supported by data, and the collection of studies from this project provide data on various types of read-aloud accommodations that states can use for increasing the accessibility of their assessments while maintaining their validity.

In discussion with state members who participated in the project, a number of ways that the results can be used to benefit states were identified. These included use with peer review, justifying states' use of specific accommodations for Title I accountability, validating the use of read-aloud accommodations with various state tests, providing access to additional data from validity studies (such as the project database), and possibly providing assistance toward standardizing the use of accommodations and obtaining a more consistent and cost-effective product from vendors.

In addition, the CTEAG project provides evidence that collaborative research can be conducted by a group of states and that all can benefit from the work. Although each state may be somewhat different in its capacity to do research, with the right kind of support and guidance a state can conduct its own validity study on test accommodations. The guidelines provided in this project can guide any state that wants to do such a study.

States also can learn from one another based on the findings that are shared from this type of collaborative project. It is feasible that a state that uses a similar type of accommodation in its assessment program, such as a read-aloud delivered by a trained individual or a read-aloud delivered by computer, can assume that the findings from one state's study will generalize to their own as long as the state has carefully considered the comparability of the constructs being measured. In other words, an accommodation that increases the accessibility of the test for students with disabilities in one state should do likewise in another state, given that other aspects of the state content standards and assessment programs are similar.

As noted earlier, states continue to require additional evidence in support of their decisions on the use of various accommodations. The results from the CTEAG project can help provide this information. Because the final reports and database include information on research designs, procedures, and findings from the five validity studies, it can be a valuable resource for every state that needs to reference any existing data and evidence to inform, evaluate, and support its decisions on the use of various types of accommodations. Also, because the results from the states cover a range of different accommodations, the project provides a resource that few, if any, states would be able to develop by themselves. This compilation and sharing of

information can be a great benefit for states in meeting NCLB peer review requirements for standards and assessments by providing additional validity information about student achievement and the performance of students with disabilities.

Teachers: Many educators can use more guidance based on findings from scientifically based research projects that will help them make informed decisions when it comes to the best ways to assess their students. Classroom teachers, in particular, will benefit by using the results from the CTEAG studies to inform their decisions about which accommodations should be used with individual students. It is important that teachers understand how certain accommodations impact the validity of the tests they give and the interpretations of scores from the tests. It is hoped that teachers also will become more sophisticated with their use of data and findings from research studies in the process of working with IEP teams to select which accommodations will be most appropriate for increasing access for individual students.

Policymakers: Policymakers should use empirical evidence when making policies related to which accommodations should be allowed on specific tests and how test scores will be interpreted if a specific accommodation is used. Policy decisions should be based on research findings like the results from the CTEAG validity studies. This empirical data can inform and support the validity of the assessment program.

Another area that is important for policymakers to recognize is their influence on and support of additional research on test validity. They can encourage more research on accommodations data from their state assessment programs through such actions as making operational test data available to researchers or collecting more information on the use of accommodations in their jurisdiction. More data on the specific types of disabilities, such as ADHD, dyslexia, or other learning disabilities, and what accommodations were provided to these students would also be useful to collect.

Researchers: As noted earlier, the research design used in the CTEAG studies can easily be reproduced by additional states in future years to study the validity of other accommodations. Other researchers also can utilize the design if they wish to conduct studies of this type. The findings can then be added to the growing body of knowledge that is being built from this project and others. It is hoped that a growing compilation of validity studies on test accommodations is created, with each new study being added to the existing collection. In this way, a shared body of knowledge will exist for all interested in accessing more information on the topic.

In addition, the project is providing data from the five studies in a database that can be used by researchers to conduct different types of analyses to answer their particular questions. More information on the database can be found in the following section.

Although a number of issues described in the previous section are challenging to the field (small samples, heterogeneity of samples, bundled accommodations, etc.), researchers should be

able to address many of these issues in the future and continuously improve the designs and approaches used to study test accommodations.

Need for Additional Research: Despite the numerous vexing issues and challenges in conducting these types of studies, as described above, much progress has been made in recent years in carrying out research on the validity of test accommodations. Increasing amounts of high-quality research are being done in this area and the findings are being shared more widely so they can have a positive impact on policy and practice. In the Appendix, a list of experimental studies and recent research on the validity of accommodations is provided (see Accommodations Bibliography).

It is critical that researchers continue doing these types of studies in spite of the many challenges and issues, as the results will only help to increase the accessibility of assessments for students with disabilities and improve the overall quality and validity of assessment results. The CTEAG project and its collaborative approach to doing a collection of state-based research studies is just a small, but significant, step forward in helping to answer the many questions pertaining to the validity of accommodations.

Information Provided in the CTEAG Technical Report and Project Database

The other two deliverables from the CTEAG project are a technical report and a project database. The technical report includes more details on each of the state studies, including specific information on the samples, test instruments, data, test score breakdowns, item statistics, factor analysis statistics, and analysis of variance results. An expanded review of the literature also is provided in this report.

The project database includes all data from the studies, at the student and item level, as well as output from the statistical analyses that were conducted for each of the state studies. Additional information on the studies is provided in the database so other researchers can replicate the analyses or do further analyses using the data.

Appendices

The following documents are included as appendices to this report:

- A. Glossary of Key Terms and Acronyms
- B. Organizational Structure and Key Personnel for the CTEAG Project
- C. Guidelines in Planning for the Accommodations Validity Studies: Steps for States to Follow in Preparing to Conduct a Study
- D. Accommodations Bibliography: Experimental Studies
- E. References

Glossary of Key Terms and Acronyms

AA-MAS – Alternate Assessment based on Modified Achievement Standards. An AA-MAS is designed to enable some students with disabilities to meaningfully participate in the statewide assessment system. A modified academic achievement standard is an expectation of performance that is challenging for eligible students, but may be less difficult than a grade-level academic achievement standard.

ANOVA – analysis of variance. A statistical procedure that determines whether the means of several groups are all equal.

Construct irrelevant variance – sources of variance associated with extraneous features of the test that are not in the intended construct that is being measured.

Differential boost – indicates that SWD receive significantly larger gains in overall test score from an accommodation than SWOD, and although both groups can receive improved test scores by using the accommodation, the gains for the SWD are differentially larger.

Factor analysis – a statistical method used to reduce a large number of variables to a smaller set of highly intercorrelated variables, or “factors.”

IEP – Individualized Education Program

NCLB – No Child Left Behind. The federal legislation for the Elementary and Secondary Education Act, signed into law in 2002.

SWD – students with disabilities

SWOD – students without disabilities

Test construct – the concept that a test intends to measure; the knowledge, skills, and abilities related to a specific area (e.g., mathematics, reading comprehension, etc.) that is being measured by a test.

Validity – the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Although classical models divided the concept into various “validities,” such as content validity, criterion validity, and construct validity, the modern view is that validity is a single unitary construct.

Organizational Structure and Key Personnel for the CTEAG Project

Connecticut State Department of Education

Barbara Beaudin, Project Director
Mohamed Dirir, Psychometrician and Project Analyst
Joe Amenta, CT State Project Lead
Gil Andrada, CT State Project Lead
Janet Struck, CT State Education Consultant

Participating States and Representatives

Connecticut: Mohamed Dirir, Joe Amenta
Kentucky: Ken Draut, Roger Ervin
Michigan: Steve Viger, Adam Wyse
Nevada: Carol Crothers, Andrew Parr

Council of Chief State School Officers (Primary Contractor)

Doug Rindone, TILSA Advisor
Sandra Warren, ASES Advisor
Adam Petermann, CCSSO Liaison

State Collaborative on Assessment and Student Standards (SCASS)

Technical Issues in Large-Scale Assessment (TILSA) Consortium
Assessing Special Education Students (ASES) Consortium

Subcontractors and Partners

John Olson, Principal Investigator and Project Coordinator
Martha Thurlow, Project Advisor and Partner
Eileen Ahearn, Project Advisor and Partner
Phoebe Winter, Project Evaluator

Guidelines in Planning for the Accommodations Validity Studies: Steps for States to Follow in Preparing to Conduct a Study

[Note: This document was shared with participating states in November 2007]

A. Building the project test forms for the study.

1. If at all possible, the participating state should use an existing form of the state test (such as a released form, alternative form, breach form, etc.). Another option is to assemble the project test forms from a pool of existing items that the state already has. If a state does not have an existing form or pool of items that are available, staff should contact the project to discuss possible ways to acquire a test for use in the study. Mohamed Dirir and John Olson will provide advice as necessary to state staff.
2. Divide the form into two parallel half-forms. The two half-forms must be as parallel as possible in the following features:
 - i) The content domains measured should be the same, or as similar as possible, for the two parts.
 - ii) The number of items from each item format (i.e., multiple-choice, open-ended), number of score points, and test length must be equal.
 - iii) The difficulties of the two half-forms must also match. This could be achieved by matching the IRT difficulty parameters and other item statistics.
3. State staff will be responsible for creating the two parallel forms. This will require the involvement of the content and psychometric staff. Project staff will provide advice upon request for building the forms.
4. The project researchers and the participating state will be responsible for the review and approval of the parallel forms.
5. The participating state will arrange production, printing, shipping/receiving, and scoring of the test with their test vendor.

B. Selecting the subject area, grade, and type of accommodations to study.

1. Each state will be testing only one grade and one subject (mathematics or reading).
2. Select the type of accommodation you want to study. Some examples of accommodations to study are: read aloud, online test readers, scribes, calculators, extended time, and variations in test setting (small group, individual). For this project, it is important to have a variety of different accommodations studied, as well as subject areas and grades, so plans will need to be discussed with project staff before the decision is final.
3. Make sure you have a large enough number of students for the type of accommodation

you want to study. The study requires a minimum of 150–200 students who are eligible (i.e., required per their IEP or 504 plan) for using the accommodation.

4. Although some students receive multiple accommodations as a normal part of their assessment, a single accommodation could be more reasonable and more straightforward to study. If possible, examine your data on accommodation use during the assessment to determine which combinations are most frequently used, and whether some accommodations are used alone by sufficient numbers of students.
5. There could be situations where students use more than one accommodation out of necessity. As an example, a student who needs a human reader would also have extended time and special setting accommodations. Normally, these three accommodations are interdependent, the latter two being necessary conditions for the main accommodation, the human reader. In such situations, you could focus on the major accommodation and assume that the secondary accommodations are part of the major accommodation.

C. Sampling students.

1. As noted above, a minimum of 150–200 students is needed for each of the two groups: the students with disabilities (SWD) group and the students without disabilities (SWOD) group.
2. Examine the SWD group in your state that is eligible for the selected accommodation first. Check the composition of this population (gender, ethnicity, SES, type of community, etc.). If the count of this group is more than 700, you may need to sample from the group due to the cost of testing.
3. After you are satisfied with your selection of the SWD group, select a sample from the SWOD population. As much as possible, this sample should match the demographics of the SWD group. It may be simpler to select this sample from the same schools where you will be testing the SWD group. The procedures used to select the SWD and SWOD samples for the study need to be reviewed by project staff before they are finalized.

D. Test administration.

1. Before the test administration, the state will need to provide written test administration instructions that are unique to this study for test examiners.
2. We would like to be as flexible as possible in determining the best schedule for conducting the study. The project management will discuss your test administration schedule and the project timelines with you in light of the state's testing cycle.
3. It might be a good idea to administer the project test forms as a pilot or supplemental test administration during the normal test administration window.

4. Since students without disabilities have never used accommodations before, you should let them practice the accommodation in a session or two about two to four weeks prior to the test administration to get them more familiar with it. It is suggested that these students use the accommodation once prior to the test administration and then again when the directions for administration are given to the students on the day of the assessment.
5. In case you are planning to use an accommodation that you have not used before, you will need to give practice time for both groups. For example, if your state is using a computer-delivered screen reader accommodation for the first time, you will need to give students a chance to become familiar with the screen reader and the technology that comes with it. As suggested before, students should be allowed to use the accommodation at least a couple of weeks prior to the administration as well as when directions are given on the assessment day. The procedures to be used for administration and practice will need to be reviewed by project staff before being finalized.

E. Data.

1. Project staff will conduct all analyses. The main data to be analyzed for the study are the student scores from the test administration. After the student responses have been scored, we will need student-level data, including raw data, scoring keys, item scores, student ID, and other information, from your test contractor.
2. We also plan to cross-reference test performance with disability categories, needs, and services for the SWD group. For that to happen, we will need you to share with us your special education data. Each student record should contain:
 - Student ID
 - Standard demographic data, such as gender, race/ethnicity, and SES
 - Information about the student's disability, such as type of disability and services
 - If the student's disability is academic, whether it is in the area of reading, math, or both
 - Teacher's judgment as to whether student is showing grade-level classroom performance (or some other indicator of student performance)
 - School information, such as name of school, grade, class
 - Item response data
 - Any other indicators you think would be useful in the study

These variables will be on the same student record as the student response variables. The project will provide you with a list of all specific data requirements and discuss with you the format and layout for the files.

3. To make it possible for us to help you create and check the development of parallel forms, we need to receive the test data that is used to build the half-forms. This includes IRT item parameters, classical item and test statistics, item objectives or domains, and other data on item characteristics you may have.
4. We will work with you and your contractor on the best way to collect and code the test data before the test administration.
5. In order to be as flexible as possible, we will accept data in almost any pre-specified format, but we will need to work out the easiest way to transfer the data.
6. We assure you that we will honor the confidentiality of your data sets and protect against any tampering or misuse.

Project Contacts:

Mohamed Dirir
860-713-6884
mohamed.dirir@ct.gov

John Olson
830-438-4475
jmlkolson@yahoo.com

Accommodations Bibliography: Experimental Studies

- Bishop-Temple, C. (2007). The effects of interactive read-alouds on the reading achievement of middle grade reading students in a core remedial program. Dissertations International, Florida Atlantic University.
- Brown, D.W. (2007). The role of reading in science: Validating graphics in large-scale science assessment. Assessment for Effective Intervention.
- Cahalan C., Mandinach E., & Camara W.J. (2002). Predictive validity of SAT I: Reasoning test for test-takers with learning disabilities and extended time accommodations. New York, NY: College Entrance Examination Board, 2002–05.
- Curtis H.A. & Kropp R.P. (1961). A comparison of scores obtained by administering a test normally and visually. Journal of Experimental Education, 29 (3), 249–260.
- DiRosa, F. (2008). The impact of testing accommodations on individual postsecondary student test outcomes. Dissertations International, Temple University.
- Elliott, S.N. (1999). Valid testing accommodations: Fundamental assumptions and methods for collecting validity evidence. Paper presented at the National Conference of the Council of Chief State Special Education Officers, Snowbird, UT.
- Elliott S.N., Kratochwill T.R.; McKeivitt B. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. Journal of School Psychology, 39, 3–24.
- Fletcher, J.M., Francis, D.J., Boudousquie, A., Copeland, Young, Kalinoski, & Vaughn (2006). Effects of Accommodations on High-Stakes Testing for Students with Reading Disabilities. Exceptional Children, 72 (2), 136–150.
- Fuchs L.S. (2000). Research Report on Math. The validity of test accommodations for students with learning disabilities. Differential item performance on reading and mathematics test as a function of test accommodations and disability status. Newark, DE: Delaware Education Research and Development Center, University of Delaware, 1, 0-0.
- Fuchs L.S. (2000). Research Report on Reading. The validity of test accommodations for students with learning disabilities. Differential item performance on reading and mathematics tests as a function of test accommodations and disability status. Newark, DE: Delaware Education Research and Development Center, University of Delaware, #2, 0-0.
- Helwig, R. & Tindal, G. (2003). An Experimental Analysis of Accommodation Decisions on Large-Scale Mathematics Tests. Exceptional Children, 69 (2), 211–225.

- Higgins, J., Russell, M., & Hoffman, T. (2005). Examining the Effect of Computer-Based Passage Presentation on Reading Test Performance. The Journal of Technology, Learning, and Assessment, 3 (4), 0-0.
- Hollenbeck K., Tindal G., & Almond P. (1998). Teachers' knowledge of accommodations as a validity issue in high-stakes testing. The Journal of Special Education, 32 (3), 175–183.
- Huesman R. (1999). The validity of ITBS reading comprehension test scores for learning disabled and non learning disabled students under extended-time conditions. Dissertation Abstracts International, 0-0.
- Huesman R. & Frisbie D. (2000). The validity of ITBS reading comprehension test scores for learning disabled and non learning disabled students under extended-time conditions. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Kamei-Hannan, C. (2008). Examining the Accessibility of a Computerized Adapted Test Using Assistive Technology. Journal of Visual Impairment & Blindness, 102 (5), 261–271.
- Keng L., McClarty K.L., & Davis L. L. (2008). Item-Level Comparative Analysis of Online and Paper Administrations of the Texas Assessment of Knowledge and Skills. Applied Measurement in Education, 21 (3), 207–226.
- Ketterlin-Geller, L.R., Yovanoff, P., & Tindal, G. (2007). Developing a New Paradigm for Conducting Research on Accommodations in Mathematics Testing. Exceptional Children, 73 (3), 331–347.
- Kettler R.J., Niebling B.C., Mroch A.A., Feldman E.S., & Newell M.L. (2003). Effects of testing accommodations on math and reading scores: An experimental analysis of the performance of fourth and eighth grade students with and without disabilities. Madison, Wisconsin: Wisconsin Center for Education Research, University of Wisconsin-Madison, 0-0.
- Kettler, R.J., Niebling, B.C., Mroch, A.A., Feldman, E.S., Newell, M.L., Elliot, S.N., Kratchowill, 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), 37–48.
- Kim, D.H., & Huynh, H. (2008). Computer-Based and Paper-and-Pencil Administration Mode Effects on a Statewide End-of-Course English Test. Educational and Psychological Measurement, 68 (4), 554–570.
- 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.

- Kosciolek S. & Ysseldyke J.E. (2000). Effects of a reading accommodation on the validity of a reading test. Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota, 28.
- Lewandowski, L. J., Lovett, B. J., Parolin, R., Gordon, M., & Coddling, R. S. (2007). Extended Time Accommodations and the Mathematics Performance of Students with and without ADHD. Journal of Psychoeducational Assessment, 25 (1), 17–28.
- Lewis D., Green D.R., & Miller L. (1999). Using differential item functioning analyses to assess the validity of testing accommodated students with disabilities. Paper presented at the national conference on large-scale assessment, Snowbird, UT.
- Lindstrom, J. H. & Gregg, N. (2007). The Role of Extended Time on the SAT for Students with Learning Disabilities and/or Attention-Deficit/Hyperactivity Disorder. Learning Disabilities Research and Practice, 22 (2), 85–95.
- Lovett, B. J. (2008). Determinants of postsecondary students' performance on timed examinations: Implications for extended time testing accommodations. Dissertations International, Syracuse University, 0-0.
- Middleton, K.V. (2007). The effect of a read-aloud accommodation on items on a reading comprehension test for students with reading-based learning disabilities. Dissertations International, The University of Iowa, 0-0.
- Robinson G. & Conway R. (1990). The effects of Irlen colored lenses on students' specific reading skills and their perception of ability: A 12-month validity study. Journal of Learning Disabilities, 23 (10), 589–596.
- Russell M. & Haney W. (1997). Testing writing on computers: an experiment comparing student performance on tests conducted via computer and via paper-and-pencil. Educational Policy Analysis Archives, 5 (3).
- Schulte A.G. (2000). Experimental analysis of the effects of testing accommodations on student's standardized mathematics test scores. Dissertation Abstracts International, 61/08 , 3056
- Schulte A.G., Elliott, S.N., & Kratochwill, T.R, (2000). Experimental analysis of the effects of testing accommodations on students' standardized mathematics test scores. Paper presented at the annual meeting of the Council of Chief State School Officers (CCSSO), Snowbird, UT.
- Thornton A.E., Reese L.M.; Pashley P.J., & Dalessandro S.P.; (2002). Predictive validity of accommodated LSAT scores. Newton, Pennsylvania: Law School Admission Council, 0-0.

- Tindal G., Heath B., Hollenbeck K., Almond P., & Harniss M. (1998). Accommodating students with disabilities on large-scale tests: An experimental study. Exceptional Children, 64 (4), 439–450.
- Weaver S.M. (1993). The validity of the use of extended and untimed testing for postsecondary students with learning disabilities. Dissertation Abstracts International, 55/03, 535.
- Weston T. (2002). The Validity of Oral Accommodation in Testing. Palo Alto, CA: NAEP Validity Studies Panel.
- Weston, T.J. (2003). The Validity of Oral Accommodation Testing: NAEP Validity Studies. Washington D.C.: National Center for Education Statistics, Working Paper No. 2003-06, April 2003.
- Woods, K. (2004). Deciding to Provide a Reader in Examinations for the General Certificate of Secondary Education (GCSE): Questions about Validity and Inclusion. British Journal of Special Education, 31 (3), 122–124.
- Zurcher R. & Bryant D.P. (2001). The validity and comparability of entrance examination scores after accommodations are made for students with LD. Journal of Learning Disabilities, 34 (5) , 462–471.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Psychological Association.
- Cahalan-Laitusis, C. & Cook, L. (eds.), (2007). *Accommodating students with disabilities on state assessments: What works?* Arlington, VA: Council for Exceptional Children.
- Dolan, R.P., Hall, T.E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning, and Assessment*, 3(7). Available from <http://www.jtla.org>.
- 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.
- 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.
- Huynh, H., & Barton, K. (2006). Performance of students with disabilities under regular and oral administrations of a high-stakes reading examination. *Applied Measurement In Education*, 19, 21–39.
- 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.
- National Research Council (1999). *Uncommon Measures: Equivalence and Linkage among Educational Tests*. Washington, DC: National Academy Press.
- O’Neil, T., Sireci, S.G., & Huff, K.L. (2003-2004). Evaluating the consistency of test content across two successive administrations of a state-mandated assessment. *Educational Assessment*, 9, 129–151.
- Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment*, 5, 299–321.
- Sireci, S., Scarpati, S., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75, 457–490.
- Thurlow, M.L. (April, 2007). *Research Impact on State Accommodation Policies for Students with Disabilities*. NCEO, University of Minnesota.

- Thurlow, M.L. (2007). State policies and accommodations: Issues and implications. In C. Cahalan-Laitusis & L. Cook (eds.), *Accommodating students with disabilities on state assessments: What works?* Arlington, VA: Council for Exceptional Children.
- 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 from www.nceo.info/OnlinePubs/Technical26.htm.
- Thurlow, M.L., Moen, R., & Altman, J. (2006). *Annual performance reports: 2003-2004 state assessment data* (available on the NCEO Web site at <http://www.nceo.info/OnlinePubs/APRsummary2006.pdf>). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- 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/Accomm/TaskComparability.htm>.
- Tippets, E., & Michaels, H. (1997, March). *Factor structure invariance of accommodated and non-accommodated performance assessment*. Paper presented at the National Council on Measurement in Education, Chicago.
- U.S. Department of Education. (April 2004). *Standards and assessment peer review guidance: Information and examples for meeting the requirements of the No Child Left Behind Act of 2001*. Washington, DC: Office of Elementary and Secondary Education.
- Zenisky, A.L. & Sireci, S.G. (2007). *A Summary of the Research of the Effects of Test Accommodations: 2005 through 2006*. Center for Educational Assessment, University of Massachusetts-Amherst.



One Massachusetts Avenue, NW, Suite 700
Washington, DC 20001-1431
phone: 202.336.7000
fax: 202.408.8072
www.ccsso.org