

Lessons Learned from State Validity Studies

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
Gilbert Andrada
Joseph Amenta
Connecticut State Department of Education

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.

Lessons Learned from State Validity Studies

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

Gilbert Andrada, Primary Author

Joseph Amenta, Primary Author

The primary purpose of the Connecticut Enhanced Assessment Grant (CTEAG) was to conduct research on the impact of test accommodations on the performance of students and to examine the characteristics of test forms and items. A multi-state collaborative conducted research from the context of several state departments of education. Unique among the grant collaborators, Connecticut conducted two related studies as part of this grant. The results and recommendations from the studies done in all collaboration states are reported in the *Technical Report 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)*.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

Steven L. Paine (West Virginia), President
Susan Gendron (Maine), Past-President

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

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

Connecticut Enhanced Assessment Grant

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

Lessons Learned from State Validity Studies

**Gilbert Andrada
Joseph Amenta
Connecticut State Department of Education
June 2010**

Purpose of the CTEAG Lessons Learned Report

The primary purpose of the Connecticut Enhanced Assessment Grant (CTEAG) was to conduct research on the impact of test accommodations on the performance of students and to examine the characteristics of test forms and items. A multi-state collaborative conducted research from the context of several state departments of education. Unique among the grant collaborators, Connecticut conducted two related studies as part of this grant. The results and recommendations from the studies done in all collaboration states are reported in the *Technical Report 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)*.

In Connecticut, the researchers wanted to observe the effects of automated read-aloud accommodation implementations using a computerized delivery system to discover ways to make them effective. This report predominantly describes the context within which the CTEAG studies were conducted in Connecticut. Summarizations of lessons learned from collaborating states were also included.

Background

The Lessons Learned from the initial Connecticut study led to enhancements that were implemented for investigation in a second study. As Connecticut continues to move toward computer delivered tests, studies like these help to inform the ways that technology can be employed to meet the needs of all students.

For many years now, it has been logistically challenging for schools to adhere to the Connecticut statewide testing policy, which required a single, certified teacher as a reader for each student with disabilities who needed a reader accommodation. Compromises to the ideal testing conditions have not always benefitted students. In many cases, the read-aloud test accommodations have been delivered by non-certified staff persons that were ostensibly supervised by certified staff persons. To the extent that this was happening, there were many disadvantages to the students.

Administering the assessment in a group setting was not consistent with the individualized nature of the accommodation. In a group setting, students were not able to proceed at their own speed. Faster students were hampered by the time the reader had to spend with students who required a slower pace. Conversely, students who required slower paces were adversely affected by the students requiring a faster pace. Additionally, the read aloud experience of students could also be affected by readers who may have heavy ethnic accents, may have inadvertently provided cues, and who may have read some items differently than others.

Consider the math problem “ $7 - 3 = ?$ ”. This could be correctly read as “seven minus three,” or incorrectly read as “seven take away three.” In the second case, a cue to the relevant mathematical operation was provided. Using a technological solution, such as text vocalization software for this and other tests offered a solution to these challenges.

The first problem that an automated read-aloud accommodation solved was the need to obtain a certified teacher for each student. Students who received this accommodation could have a testing experience that more closely approximated that of the students who used the standard administration of a statewide test. One proctor, in a computer lab setting, could monitor the entire test administration. Next, students using this technology could proceed at their own pace because they would have their own test delivered by a computer that they controlled. Furthermore, this type of automation would employ instances of a single, standardized, vocalization of the test materials – instructions, directions, and items. This reading could be vetted by Connecticut State Department of Education student assessment professionals for appropriate criteria, including pronunciation, pace, emphasis, and all other issues of vocalized delivery. As testing continues to move toward the computerized arena and a testing world where statewide tests may be delivered to all students on computer, the use of vocalized text may also be made more widely available, thus eliminating the moniker of accommodation for this technology.

Connecticut Study #1: Vocalized Text on a Test of Reading Comprehension

In the spring of 2008 the first of two validity studies were conducted as part of the CTEAG. In the process of conducting those studies, several useful observations surfaced in addition to the results of the studies. These additional findings included the effect of accommodations on the psychometric characteristics of the items, dimensionality of the tests, and the extent to which the accommodation appeared to accommodate disability. In our additional findings from the administration of a computer-delivered read-aloud accommodation, voicing quality, vocal cues, and the need for visual cues were revealed in the first study as important factors affecting the quality of the testing experience.

Voicing quality

There were two options considered for producing a vocalized presentation of text materials -- automated text readers such as the TextHelp™ application, and recorded human readers. There was no need for on-the-fly text vocalization because the studies used standardized forms. The quality of the automated technology, however, was not found to be satisfactory. The default values were used in this first study, but the automated product provided by TextHelp™ was customizable for pacing, pronunciation, order and emphasis in a post-production environment prior to release. This work required a forms review to check the fidelity of the voicing to the text.

A recorded human voice would have required some similar effort. In our initial use of the new automated reader system, some of the issues we considered included:

Issues:

- Standardized vocal inflections for
 - Parenthetical text
 - Italicized text
 - Bolded or highlighted text
 - Bulleted Lists
 - Underlined text
 - Superscript text meant to indicate a footnote
- Standardized pause lengths for
 - Periods
 - Commas
 - Semicolons
 - Colons
 - Double dashes
 - Ellipsis

- Trouble with words that have more than one pronunciation (lead, read, live, present, winds, tear, St., bows, etc.)

Vocal cues

The state consortium considered the need for a way to signal the reading of different kinds of text found on an assessment. For example, should there be a different voice or voice characterization for different parts of a test?

- Directions
- Item stems
- Response choices
- Passages

Visual cues

In Connecticut, standardized paper and pencil tests use consumable forms and do not have the option bullets (a., b., c., etc) alongside multiple choice response choices. This convention would provide a visual cue for what is being read if the text were vocalized. Furthermore word-for-word highlighting as text was being read would be a preferred feature in a vocalized text environment so that test takers can more easily track what is being read or what part of the visible screen is being vocalized.

Study #1: Survey Results

The students participating in the text reader accommodation study were asked to complete an optional survey about their experience. The survey consisted of two main sections where students were asked to rate statements related to online testing and to the use of the text reader accommodation on a 1-5 Likert scale. This survey was created to allow students to anonymously offer their impressions. Of the approximately 460 students participating in this study, 160 students submitted responses to the survey. The results are presented in the table on page 6.

According to the survey, about 75% of the students had never used a text reader prior to this study. In general, results indicate a positive experience with both the online testing format and the text reader accommodation. When asked if they would like to continue to use the text reader function as part of their regular classes, approximately 67% of the students agreed or strongly agreed. The students found navigating in an online testing environment to be both easy and helpful in terms of concentrating on the reading passages while listening to the vocalized text.

Question	Strongly disagree	Disagree	Somewhat agree	Agree	Strongly agree
1. The computer testing system was easy to use.	3.8% (6)	5.0% (8)	16.3% (26)	39.4% (63)	35.6% (57)
2. The reading test on-line was easier to take than taking a paper and pencil test.	3.2% (5)	6.3% (10)	13.9% (22)	31.0% (49)	45.6% (72)
3. I found it easier to concentrate on the on-line tests than a paper and pencil test	8.8% (14)	11.9% (19)	16.4% (26)	30.8% (49)	32.1% (51)
4. I found it easy to use the text reader.	6.3% (10)	5.0% (8)	17.5% (28)	36.9% (59)	34.4% (55)
5. I found it difficult to concentrate when listening to the text reader.	36.3% (57)	29.3% (46)	13.4% (21)	12.7% (20)	8.3% (13)
6. I would like to use text reader in my classes.	3.8% (6)	11.9% (19)	15.7% (25)	27.0% (43)	41.5% (66)

The survey also offered students an opportunity to submit a comment regarding their testing experience in this study. Of the 160 students surveyed, 53 students offered comments, and about 77% of these comments were positive. Here are some of positive comments made by students:

“i think that this is a very good scource to use it is alot easer and takes alot of stress off knowing that you dont have to read alot and mess up words so i would totally recomend this”

“I like text reder it was way better than paper and pencil also it was easy to use”

“With the head phones on it made it eaisier to concentrate and it was also eaiser to just click on the bubble then filling it in. I would LOVE to use this in my classes and I'm sure lots of other people would like to use them as well!!”

The negative comments made by student mostly involved the quality of the voice on the text reader section of the test:

“The level of the voice kept changing and was hard to follow at some points.”

“The voice quality isn't to great”

We also received a very positive comment sent to us as an email by one of the participating districts.

“Thanks for allowing us to participate. Many of the students enjoyed the text reader and asked if it would be used for the CMT, because they thought students would perform better with it. I told them that’s why we’re taking the pilot test, so we’ll see what the data shows.”

On behalf of the students and faculty here, thanks again for the opportunity!”

Since this survey was both optional and anonymous, there was no opportunity to compare student responses to any other data collected. For example, there was no way to determine if students with disabilities perceived this accommodation differently than those without.

Proposed Considerations for Study #2

Nearing the completion of the CTEAG Study #1, we found that we had discovered many things about vocalized text planning, production, and delivery of the computerized read-aloud test accommodation that warranted further study and development. To address these research and evaluation questions, a second study under the auspices of the CTEAG was proposed.

The findings and observations from CTEAG Study #1 were considered in developing CTEAG Study #2 and the following research questions were developed:

Vocal cues

Should there be a different voice or characterization for different types of text?
(Human vs. Digital voice)

- Directions
- Item stems
- Response choices

Visual cues

- Should students have individual sentences/words highlighted as they are being read?

Research Questions for new study:

1. Can we expand the use of text readers to accommodate a math assessment?
2. Does the accommodation change the item properties?
3. Does the accommodation change the construct of math assessment?
4. Does the accommodation increase performance of students with disabilities more than performance of students without disabilities?
Note that questions 2-4 are the same as those asked for study #1.

Address concerns from the previous study:

- Quality of voice during text-reader portion
- Visual cues provided on computer screen during text-reader portion

Evaluation Question:

Can text readers replace human readers in large scale mathematics assessments?

1. What are the logistical barriers and advantages?
2. What are the technological barriers and advantages?
3. Should we consider this as a permanent feature for all subjects as part of a computerized, statewide, large-scale assessment?
4. Can we standardize the manner in which read-aloud material is delivered to students who need this accommodation or, if policy permits, any student who would want to use this feature?

**Connecticut Study #2:
Vocalized Text on a Test of Mathematics**

In our second CTEAG study, conducted in the spring of 2009, we were able to directly address many of the concerns that became apparent in our first study. Since the testing vendor we used remained the same, this provided us the opportunity to move quickly in creating a computer delivered testing environment with the vocalized text in a manner we felt better suited students' needs.

Addressing Lessons Learned from Study #1

To address the vocal quality issue in Study #1, a recorded human voice rather than a digital one was used. This required that we provide our test developers a script of the tests indicating what should be read and in some specific cases how certain words should be pronounced. Our math tests also contained charts and graphs that we wanted to have vocalized and therefore required a consistent and logical approach. We chose one female voice for all vocalized text.

Visually, to help cue students, individual item stems and response choices were highlighted as they were being read to students. Therefore a student could choose a particular part of a test question to have read as often as needed. However, the consortium decided not to have individual words highlighted as they were being read. This decision was based on both research information and practical constraints.

Study #2: Survey Results

All students who participate in this study were also asked to complete an online survey. Unlike the first survey, this survey tracked the respondents; therefore results were able to be disaggregated. Survey #2 shows results by gender and students with and without disabilities.

In general, most students seemed to have a positive attitude toward a computer delivered test. The survey showed 84% of students with disabilities and 76% of students without disabilities agreeing or strongly agreeing that a computer math test seems easier to take than a paper and pencil test. In terms of

having a vocalized text function, 75% of both groups agreed or strongly agreed as to its ease of use. It is interesting to note, when asked if it is easy to think when listening to text read to them, 74% of students with disabilities but only 58% of students without disabilities agreed or strongly agreed.

Study #2 Survey Results

1. Have you ever used a text reader before this study?

	Students with Disabilities	Students without Disabilities	Male	Female
YES	33.8%	19.0%	28.7%	21.6%
NO	66.2%	81.0%	71.3%	78.4%

2. The computer test was easy to use.

	Students with Disabilities	Students without Disabilities	Male	Female
Strongly Agree	48.7%	48.4%	47.1%	50.2%
Agree	33.5%	36.8%	36.3%	34.4%
Somewhat Agree	14.1%	12.2%	13.0%	13.1%
Disagree	1.1%	1.4%	0.6%	2.1%
Strongly Disagree	2.6%	1.1%	3.0%	0.3%

3. The math test on the computer was easier to take than a paper and pencil test.

	Students with Disabilities	Students without Disabilities	Male	Female
Strongly Agree	58.0%	53.3%	59.8%	50.2%
Agree	26.0%	22.7%	22.4%	26.1%
Somewhat Agree	10.0%	16.7%	11.8%	16.2%
Disagree	3.0%	4.2%	2.7%	4.8%
Strongly Disagree	3.0%	3.1%	3.3%	2.7%

4. I found it easier to think on the computer test than a paper and pencil test.

	Students with Disabilities	Students without Disabilities	Male	Female
Strongly Agree	48.0%	43.1%	46.5%	43.6%
Agree	27.1%	22.1%	25.7%	22.7%
Somewhat Agree	12.3%	20.1%	14.5%	19.2%
Disagree	7.1%	10.2%	9.4%	8.2%
Strongly Disagree	5.6%	4.5%	3.9%	6.2%

5. I found it easy to use the text reader.

	Students with Disabilities	Students without Disabilities	Male	Female
Strongly Agree	49.8%	43.6%	49.8%	42.3%
Agree	26.4%	34.0%	28.7%	33.0%
Somewhat Agree	13.0%	12.5%	10.6%	15.1%
Disagree	5.6%	6.2%	5.4%	6.5%
Strongly Disagree	5.2%	3.7%	5.4%	3.1%

6. I found it easy to think when listening to the computer text reader.

	Students with Disabilities	Students without Disabilities	Male	Female
Strongly Agree	51.3%	35.1%	42.9%	41.2%
Agree	22.3%	23.2%	24.5%	21.0%
Somewhat Agree	17.1%	20.1%	19.0%	18.6%
Disagree	4.5%	11.0%	6.6%	10.0%
Strongly Disagree	4.8%	10.5%	6.9%	9.3%

7. I would like to use text reader in my classes.

	Students with Disabilities	Students without Disabilities	Male	Female
Strongly Agree	46.1%	46.5%	45.0%	47.8%
Agree	20.4%	24.9%	23.0%	23.0%
Somewhat Agree	17.1%	16.1%	16.3%	16.8%
Disagree	11.2%	6.8%	11.2%	5.8%
Strongly Disagree	5.2%	5.7%	4.5%	6.5%

8. I liked the voice used during the text reader.

	Students with Disabilities	Students without Disabilities	Male	Female
Strongly Agree	34.9%	28.9%	30.5%	32.6%
Agree	32.3%	27.5%	28.7%	30.6%
Somewhat Agree	16.4%	24.1%	21.5%	19.9%
Disagree	8.2%	8.5%	8.8%	7.9%
Strongly Disagree	8.2%	11.0%	10.6%	8.9%

9. It was easy to have words read to me.

	Students with Disabilities	Students without Disabilities	Male	Female
Strongly Agree	61.7%	48.7%	54.7%	54.0%
Agree	24.9%	29.7%	28.4%	26.8%
Somewhat Agree	5.2%	9.9%	6.0%	10.0%
Disagree	5.2%	8.8%	7.6%	6.9%
Strongly Disagree	3.0%	2.8%	3.3%	2.4%

10. I liked it when sentences and words were highlighted for me.

	Students with Disabilities	Students without Disabilities	Male	Female
Strongly Agree	44.2%	34.3%	38.7%	38.5%
Agree	29.4%	34.3%	31.7%	32.6%
Somewhat Agree	14.1%	20.1%	16.0%	19.2%
Disagree	5.9%	7.4%	7.3%	6.2%
Strongly Disagree	6.3%	4.0%	6.3%	3.4%

Implications for Connecticut

Other than the research questions we needed answered through these two studies, there were practical matters that also needed to be addressed. The adoption of a computer based test as a means of delivering an accommodation such as vocalized text involves a much larger vision of how statewide testing in Connecticut can evolve. In 2010 we will be providing a computer based test to more students with disabilities than ever before. We have also started to offer a partial vocalized text function in most of our tests. These decisions were made as a by-product of how we were able to prepare our schools and our testing vendor for the vision we had as we discussed, developed and implemented these two studies.

Lessons Learned Across the Two Connecticut Studies

- Computer based assessment with vocalized text may involve programming limitations and time restraints; determine which navigational and function features are essential.
- Availability of functioning hardware i.e. working computers, adequate internet access and headphones are as important as the computer based assessment format.
- Adequate time must be allotted to prepare both testing proctors and students on how to navigate through an online assessment environment with accommodations.
- There is nothing more valuable than the feedback from a field test. Listen to the feedback and be prepared to make adjustments accordingly.

Appendix

Summaries of Lessons Learned Reported by State Researchers from Other Validity Studies in the CTEAG Project

In addition to Connecticut, three other states participated in the CTEAG project: Kentucky, Michigan, and Nevada. Key representatives in each state were asked to provide brief comments on the lessons learned in their state from conducting their validity study. The format for these comments was not pre-defined and the information they submitted could address any issues they thought to be important. The intent was to get their unbiased and objective perspectives on how the process went, any special considerations that were unique to their state, and anything else they wanted to include. The following input from the states reveal a variety of aspects on the studies as well as various testing issues that should be considered.

Lessons Learned by Kentucky

Kentucky chose to study the effect of read aloud accommodation on test scores for students with disabilities (SWD) in grade 4. The validity study findings showed increased test scores for SWD more than for students without disabilities (SWOD) and the difference in score gain due to the read aloud accommodation between the two groups was significant.

- Kentucky would suggest conducting future studies in a different time of the school year, probably in the fall. In 2009, Kentucky experienced a significant number of lost instruction days related to inclement weather and had three federal disaster declarations for weather during the 2008-2009 school year. While schools were originally willing to participate in the study, priority shifted from the study to administering the state summative assessment for state accountability and No Child Left Behind results. Thus, fewer SWD and SWOD students participated than in the original plan. In addition, although financial incentives were offered to schools to participate in the study, they came second to schools' desire to spend time on student instruction instead of testing.

Were this study conducted again, Kentucky recommends:

- Provisions for on-site monitoring of the assessment process to gain an understanding of how accommodations are administered at the individual student level.
- Expanding the student read-aloud accommodation testing to students at middle school and high school levels to measure age affect.

Kentucky has benefited from participating in the CTEAG study. With states receiving new direction from the United States Department of Education, Kentucky has utilized these report findings to support Kentucky's application for Race to the Top and State Fiscal Stabilization Funds Phase 2.

Lessons Learned by Michigan

It is critical to remember that students eligible for the 2% assessments are perhaps one of the more heterogeneous groups of students eligible for alternate assessments. To this end, states need to be mindful of their approaches to item modification and try to avoid "cookie cutter" or one size fits all approaches. Granted, such approaches are commonly done when creating the typical accommodated forms for general assessments. While some students' performance is enhanced by pointing out key topics using bold print, underlining topic sentences in paragraphs, chunking paragraphs, outlining items in boxes and the many other enhancements gleaned from solid educational psychology research, we must remember that precisely what aids one student could greatly diminish another's capacity to decode the intended information and respond to an item or prompt.

Students at a younger age are more likely to be able to assimilate scaffolding information and enhanced directions into their current schema because it truly is not as developed and "set in stone." However, the more the student matures, the more metacognitively aware they are likely to be. Therefore, if there is a noticeable difference between how they are used to interacting with an item and how they interact with the item in the face of other information surrounding the prompt, the students will experience greater amounts of cognitive dissonance. This, of course, will increase the likelihood that a mismatch between what is presented and expectations will force the student into an accommodation rather than an assimilation situation as they attempt to make sense of the task at hand.

Finally, the overall expectation is that whatever is presented on the assessment represents a situation the students should be familiar enough with that they can devote their attention to answering the prompt rather than having to filter through too much extraneous information. That is, what is assessed should reflect what they have had access to via direct instruction from the teacher. When modifying assessment items for the target population, we should probably rely more on the principles of authentic assessment born out of industrial/organizational psychology than principles of general assessment development.

Lessons Learned by Nevada

When recruiting schools for participation, the Nevada Department of Education (NDE) found:

- Initially, schools were somewhat reluctant to participate.
- Financial incentives helped a few schools agree to participate.
- The promise of receiving data that could be used to identify learning gaps for students was the more effective incentive.

Schools were allowed to choose their testing date within a broad testing window (Nov 3-Dec 19), which may have been too broad. However, no particular problems were identified.

It is possible that test administration was not standardized among all schools that participated in the study:

- Some administrations were read by a teacher, assistant, other proctor (uniformity not assumed).
- Random assignments to groups was requested but could not be assured.
- The low-stakes nature of the assessment perhaps reduced motivation in students to do well.
- There was no assurance that standard administration protocols were adhered to in all schools. The NDE assumes that standard CRT administration protocols were followed.
- The assessment was not timed – additional time, if needed to finish, was probably not provided to all students.

It was crucial to have Dr. Olson and Dr. Dirir available as the principal contacts for the validity study.